NASA SP-305

PHYSICS OF THE SPACE ENVIRONMENT

A colloquium held at
THE UNIVERSITY OF ALABAMA
Huntsville, Ala.
1970–1971

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

NASA SP-305

PHYSICS OF THE SPACE ENVIRONMENT

A colloquium held at the University of Alabama Huntsville, Ala., 1970–1971

Edited by

R. E. Smith and S. T. Wu

Scientific and Technical Information Office 1972
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
Washington, D.C.

For sale by the National Technical Information Service
Springfield, Virginia 22151 — Price $3.00

TABLE OF CONTENTS

		Page
	INTRODUCTION	1
	ACKNOWLEDGMENT	2
Chapter I. –	TRANSIENT HEATING IN THE UPPER ATMOSPHERE	3
	A. Abstract	3
	B. Introduction	3
	C. Mathematical Formulation of the Problem	4
	D. Analytic Solutions for Temperature	11
	E. Application to Impulsive Heating of the Upper Atmosphere	15
	F. Summary and Conclusions	20
	G. References	22
Chapter II.	DISSIPATION OF ELECTRIC FIELDS IN THE IONOSPHERE	25
	A. Abstract	25
	B. Introduction	25
	C. Theory	26
	D. Joule Heating – Case a: $t_E \ll T_D$	31
	E. Winds and Possible Pressure Gradients – Case b: $t_E \gg T_D$	35
	F. The Sq System	36
	G. Conclusions	37
	H. References	38
Chapter III.	STRUCTURE AND VARIATIONS OF THE HETEROSPHERE	41
	A. Abstract	41
	B. Discussion	41
Chapter IV.	THE STABLE AURORAL RED ARC OF OCTOBER 31/NOVEMBER 1, 1968 AND ITS INTERACTION WITH THE NEUTRAL ATMOSPHERE	45
	A. Abstract	45
	B. Introduction	45
	C. Summary of Observed Features	46
	D. Excitation Mechanisms	48

TABLE OF CONTENTS (Continued)

		Page
	E. Observations	49
	F. Analysis	53
	G. Plasma Pause Position	55
	H. Atmospheric Response	58
	I. λ6300 Doppler Temperature Measurements	61
	J. λ6300 Doppler Wind Measurements	62
	K. Conclusions	64
	L. References	65
	M. Bibliography	69
Chapter V.	SOME ASPECTS OF ATMOSPHERIC WAVES IN REALISTIC ATMOSPHERES	71
	A. Abstract	71
	B. Introduction	71
	C. Internal Gravity Waves	72
	D. Application to Atmospheric Tides	79
	E. Diurinal Oscillation in the Thermosphere	82
	F. Semidiurnal Oscillation in the Thermosphere	84
	G. Remaining Problems and Possible Solutions	87
	H. References	89
Chapter VI.	UPPER ATMOSPHERIC SPECTROSCOPY	91
	A. Abstract	91
	B. Introduction	91
	C. Spectral Emissions Stimulated by Atomic Particles	92
	D. Energy Levels and Cross Sections	95
	E. Dissociative Excitation	104
	F. Thermal Excitation	108
	G. Efficiencies	109
	H. The Problem of $O_2(a^1\Delta_g)$	113
	I. Discussion and Conclusion	119
	J. References	121
	K. Bibliography	127

TABLE OF CONTENTS (Concluded)

		Page
Chapter VII.	THE COMPOSITION OF THE UPPER ATMOSPHERE	129
	A. Abstract	129
	B. Introduction	129
	C. Instrumentation	130
	D. Data	131
	E. Analysis	136
	F. Helium	136
	G. Atomic Oxygen	140
	H. References	143
	I. Bibliography	145
Chapter VIII.	CHEMISTRY OF THE UPPER ATMOSPHERE AND THE ROLE OF TRANSPORT PROCESSES	147
	A. Abstract	147
	B. Introduction	147
	C. Dissociation Mechanisms	148
	D. Oxygen Reactions	150
	E. Hydrogen-Oxygen Chemistry	151
	F. Nitrogen-Oxygen Chemistry	151
	G. Transport Processes	152
	H. References	155
Chapter IX.	THE FLARE FORECASTING FIASCO – IS EVERYBODY LOOKING AT THE SAME SUN?	157
	A. Abstract	157
	B. Introduction	157
	C. The Development of the Active Region, the Flare, and Its Effects	159
	D. Forecasting Problems	161
	E. Solutions	165
	F. Summary	171

PHYSICS OF THE SPACE ENVIRONMENT

INTRODUCTION

By

R. E. Smith
Aero-Astrodynamics Laboratory
NASA-Marshall Space Flight Center,
Marshall Space Flight Center, Alabama
and
S. T. Wu
The University of Alabama in Huntsville, Huntsville, Alabama

Accurate, up-to-date information on current and predicted values of space environment parameters is essential for the design and the operation of the Space Shuttle, Saturn V, Skylab, Space Station, and future space programs. Therefore, a series of lectures concerning the present state-of-the-art of the physics of the space environment was organized and presented at The University of Alabama in Huntsville during the academic year 1970-1971 under the sponsorship of the Aerospace Environment Division, Aero-Astrodynamics Laboratory, NASA-George C. Marshall Space Flight Center.

This series of lectures contained three major topics:

1. Dynamics and transient state of the upper atmosphere.

There were five lectures relating to this topic. Dr. Gary E. Thomas, University of Colorado, gave a lecture concerning the time-dependent response of the upper atmosphere to transient heating. He showed that atmospheric disturbances resulting from geomagnetic storms are much larger in the auroral zones and that joule dissipation of auroral currents causes atmospheric variations that perturb satellite orbits. Dr. Keith D. Cole, La Trobe University, Australia, reviewed some likely sources of electric currents and fields in the auroral ionosphere and showed that two major effects of the dissipation of these currents are (a) increases in scale heights and temperatures at the site of the auroral electrojet at altitudes above approximately 100 km and (b) the generation of high wind speeds in the auroral atmosphere. Dr. L. G. Jacchia, Smithsonian Astrophysical Observatory, summarized some of the more recent findings concerning the structure of and variations in the heterosphere. Dr. R. G. Roble, National Center for Atmospheric Research, presented a detailed discussion of the theoretical and observational results for a stable auroral red arc. Finally, Dr. Richard S. Lindzen, University of Chicago, discussed the theoretical behavior of tides and gravity waves in a realistic atmosphere including the effects of viscosity, thermal conductivity, and mean shears. Special attention was given to the semidiurnal tide and to atmospheric free oscillations.

2. Chemical Composition of the upper atmosphere.

There were three lectures relating to this topic. Prof. A. E. S. Green, University of Florida, reviewed thoroughly the current status of the upper atmospheric spectroscopy and specifically illustrated the interdependence of atmospheric processes in a discussion of photon, electron, and proton stimulation of day glow and auroral emissions. Prof. Alfred O. C. Nier, University of Minnesota, discussed the neutral ionic composition of the earth's atmosphere above 100 km from the most recent results obtained by mass spectrometers carried on sounding rockets and satellites. He concluded that above 100 km, neutral N_2, O_2, O, and Ar are in diffusive equilibrium. Finally, Prof. Richard A. Craig, Florida State University, discussed the important chemical reactions affecting the neutral composition of the upper stratosphere, mesosphere, and lower thermosphere including the influences of the atmospheric transport processes.

3. Solar flare forecasting.

Last, but certainly not least, Dr. Frederick W. Ward, Jr., Air Force Cambridge Research Laboratory, discussed solar flare forecasting. He summarized the presently available techniques and equipment used in making solar observations and criticized the accuracy of the archived scientific data. He then concluded that there are at least two suns up there.

ACKNOWLEDGMENT

We would like to thank Mr. William W. Vaughan and Dr. L. L. DeVries of the Aerospace Environment Division, Aero-Astrodynamics Laboratory, NASA-George C. Marshall Space Flight Center for their support, encouragement, and suggestions.

CHAPTER I. TRANSIENT HEATING IN THE UPPER ATMOSPHERE

By

Gary E. Thomas
Department of Astro-Geophysics and Laboratory for Atmospheric and Space Physics
University of Colorado, Boulder, Colorado

A. Abstract

The time-dependent response of the upper atmosphere to transient heat sources is considered. The basic problem is that of heating a compressible, heat-conducting fluid, which is described in the one-dimensional case by an analytic solution. Comparisons with satellite drag data of such first-order solutions are shown to be useful in determining energy requirements and in determining some constraints on the spatial distribution of the heating. Recent OGO-6 mass spectrometer and interferometric temperature measurements show that atmospheric disturbances during geomagnetic storms are much more prominent in the auroral zones. These results suggest that joule dissipation of auroral currents are important contributors to orbital perturbations of satellites.

B. Introduction

The theoretical basis for the study of the thermal and physical structure of the upper atmosphere began with the classic papers of Spitzer [I-1] and Bates [I-2]. In these papers, nearly all the important heating and cooling processes which are now known to occur in the thermosphere were described. These papers appeared at a time when our knowledge of upper atmospheric properties was extremely meager, arising entirely from a few rocket measurements and ground-based ionosonde measurements. Today, we have a very complete picture of the gross properties of the atmosphere above the F2 peak as a result of satellite drag studies, largely by the two groups headed by Jacchia and King-Hele. In addition, we are now measuring the fine structure of the atmosphere by in situ probes on rockets and satellites and by methods of ground-based incoherent backscatter.

Unfortunately, our ability to theoretically describe the upper atmospheric structure and its variations in time and space has not advanced with the same degree of success. In fact, until recently our theoretical analyses of even the regular variations have largely followed the pattern set by Bates [I-2] and by the pioneering efforts of Harris and Priester

[I-3]. This pattern consists of a one-dimensional, time-dependent model in which the atmosphere is heated by solar photoionization and subsequent recombination and is cooled by molecular heat conduction and far infrared emission from atomic oxygen. This type of calculation underlies the vertical structure of the COSPAR International Reference Atmosphere (CIRA) models [I-4] which are still in extensive use. We have long realized the inadequacies of these one-dimensional models but have only recently advanced to the point of being able to model the much more complex two- and three-dimensional transfer processes in the atmosphere.

One theoretical understanding of the important short-period variations in neutral density, and presumably temperature, that are known to occur during geomagnetic storms is in an even more rudimentary state. No serious attempts to model impulsive density changes, even in a one-dimensional framework, were made until the author (with B. K. Ching) recently proposed such a solution [I-5].

This paper will describe the details of these solutions in the belief that their analytic form will assist the reader in gaining physical insight into some of the important atmospheric heat flow processes. The shortcomings of the one-dimensional description will be stressed and warnings issued against making too detailed a comparison of its predictions with the observed features. In closing, some recent advances in this area will be described.

C. Mathematical Formulation of the Problem

Because of the absence of radiatively active molecules in the earth's upper atmosphere, a dominant mode of removing heat (which is deposited by photoionization processes) is by molecular heat conduction. In addition, it is now known that mechanical modes of heat removal are important; examples of these are the generation of internal gravity waves and the effects of large scale circulation of the neutral gas. The importance of these mechanical modes arises because of the effects of strong horizontal gradients in atmospheric properties produced by heating gradients. However, if the heating occurs uniformly over horizontal scales of the order of 1000 km, it is permissible to ignore horizontal gradients. In this case, the heat flow is predominantly in the vertical direction, and we may write the one-dimensional conservation equations of mass, momentum, and energy:

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial z}(\rho W) = 0 \quad , \tag{1}$$

$$\frac{DW}{Dt} + \frac{1}{\rho}\frac{\partial p}{\partial z} + g = 0 \quad , \tag{2}$$

and

$$\rho\, c_p \frac{DT}{Dt} - \frac{Dp}{Dt} = Q \quad , \tag{3}$$

where z is the geometric altitude, t is time, ρ is mass density, W is the vertical flow velocity of the gas, p is pressure, g is the gravitational acceleration, T is the temperature, c_p is the specific heat at constant pressure, and Q is the net heating per unit volume. D/Dt denotes the convective time derivative $\partial/\partial t + W\partial/\partial z$. The earth's curvature and the altitude variation of g are ignored. The force of the ions on the neutrals (ion-neutral drag) during veritical expansion or contraction is ignored. To complete the set of equations, we assume the applicability of the perfect gas law,

$$p = \rho\, RT/M \quad , \tag{4}$$

where R is the gas constant for the air mixture and M is the mean molecular weight.

Equations (1) through (4) describe the density, pressure, temperature, and vertical velocity of a stratified gas in response to an arbitrary net heat source Q(z,t). The atmosphere is considered to be a compressible fluid of variable mean molecular weight. Under the influence of a net heating (cooling), the fluid will expand (contract); i.e., the air parcels of equal pressure will respond to the imbalance of gravitational and pressure gradient forces. They will be accelerated in the vertical direction to offset the imbalance. This acceleration [given by DW/Dt in equation (2)] is very small compared to the gravitational acceleration (shown in the appendix of Reference I-5) and may be ignored. The assumption is made that a constant-pressure parcel maintains a constant mean molecular weight as it flows vertically (hence the lack of a subscript on W that would label the velocity of the particular molecular constituent). The accuracy of this assumption has not been critically examined; however, it is likely to be better than the accuracy of the fundamental assumption of solely vertical flow.

On ignoring the initial term DW/Dt, equation (2) may be directly integrated:

$$p(z,t) = p_o \exp\left[- \int_{z_o}^{z} dz'/H(z',t) \right] \quad , \tag{5}$$

where p_o is the pressure at the lower boundary located at a fixed altitude z_o and H is the instantaneous pressure scale height RT/mg. Here, the ideal gas law equation [equation (4)] has been used. The density is given by

$$\rho = (\rho_o H_o/H)\exp\left[-\int_{z_o}^{z} dz'/H\right] \quad . \tag{6}$$

The subscript "o" refers to the quantity at altitude z_o. If molecular diffusion processes are much more rapid than turbulent processes above z_o, we may express the number density of each constituent as

$$n_i = (n_{io}\, T_o/T)\exp\left[-\int_{z_o}^{z} dz'/H_i\right] \quad , \tag{7}$$

where $H_i = kT/M_i g$ and M_i is the molecular mass of the ith constituent. The mean molecular weight is defined by

$$M = \sum M_i p_i/p \quad , \tag{8}$$

where p_i is the partial pressure given by

$$p_i = n_i\, RT/M_i \quad . \tag{9}$$

With the use of equation (6) and the continuity equation [equation (1)], we can show that the vertical velocity is given by

$$W = H\int_{z_o}^{z} \frac{dz}{H^2}\,\frac{\partial H}{\partial t} \quad . \tag{10}$$

It is clear that, provided we can describe the temperature profile T(z, t) at each instant, the remaining variables p, ρ, and W are immediately determined by the previous equations. Thus, a solution of the heat balance equation [equation (4)] for T for a given Q is the key to the solution of the problem. However, the appearance of the density in equation (4) would seem to cause an unfortunate coupling of equations (2) and (3). Fortunately, this coupling may be largely removed by a transformation to a new coordinate system provided the net heating may be easily expressed in the new coordinate system (the latter is subsequently shown to be true). We now describe the various terms contributing to the net heating Q:

1. MOLECULAR CONDUCTION

Molecular conduction is given by

$$Q_c = \frac{\partial}{\partial z}\left(K \frac{\partial T}{\partial z}\right) \quad . \tag{11}$$

Q_c is the divergence of the conductive heat flux $K(\partial T/\partial z)$, where K is the coefficient of heat conduction, given approximately by

$$K = \sum_i A_i n_i T^{s_i}/n \quad , \tag{12}$$

where n is the sum of the individual densities n_i, and A_i and s_i are empirical constants for each constituent.

2. SOLAR PHOTOIONIZATION HEATING

This is given by

$$Q_s = \rho \int d\lambda \, \eta_\lambda \, \sigma_\lambda \, F_\lambda \exp\left[-\sigma_\lambda \int_z^\infty \rho \, Ch\chi \, dz'\right] , \tag{13}$$

where η_λ is the heating efficiency of the solar radiative flux F_λ in the wavelength interval λ to $\lambda + d\lambda$, σ_λ is the cross section per unit mass, and $Ch\chi$ is the Chapman function for the solar zenith angle χ. The quantity σ_λ is an average cross section that is evaluated at the height of maximum energy deposition,

$$\sigma_\lambda = \sum_i n_i k_i/\rho \quad , \tag{14}$$

where k_i is the absorption cross sections (in square centimeters) of the ith constituent.

3. RADIATIVE COOLING

The thermal emission from the ^3P levels of atomic oxygen is given by

$$Q_r = C \, n_1 \, f_c \, (T) \quad , \tag{15}$$

where C is the rate at which energy is emitted from the $^3P_1$ and $^2P_0$ levels of atomic oxygen (whose density is denoted by n_1) and $f_c(T)$ is a quantity that determines the population of this level in a Boltzmann distribution and is only a weak function of temperature in the range of 700 to 2000°K. It is convenient to assign an average to the quantity $f_c(T)$; that is, to assume that the radiative cooling is proportional only to the atomic oxygen density [I-6].

4. IMPULSIVE HEAT SOURCES

During geomagnetic storms, the upper atmosphere receives energy input Q_g in addition to direct solar heating. Since the physics of these processes is exceedingly complex, we will parameterize this quantity assuming that it has some simple analytic form in both space (z) and time (t). The time-dependence that has been considered is given below:

$$f_1(t) = \delta(t) \tag{16}$$

$$f_2(t) = \frac{t}{\tau^2} \exp(-t/\tau) \tag{17}$$

We now consider the transformation that uncouples equations (2) and (3). We define the independent coordinate

$$\xi = \exp\left[-\int_{z_o}^{z} dz/2H\right] = \left(p/p_o\right)^{1/2} \quad . \tag{18}$$

Let us replace z with ξ and use the properties

$$\frac{\partial}{\partial z} = -\left(\frac{\xi}{2H}\right) \frac{\partial}{\partial \xi} \quad , \tag{19}$$

$$\frac{\partial T}{\partial z} = \left.\frac{\partial T}{\partial \xi}\right|_{\xi} + \frac{\partial T}{\partial \xi} \left.\frac{\partial \xi}{\partial t}\right|_{z} \quad , \tag{20}$$

and

$$\left.\frac{\partial \xi}{\partial t}\right|_{z} = \frac{\xi W}{H} \quad , \tag{21}$$

noting that $|_{\xi}$ and $|_{z}$ indicate that the variables ξ and z are held fixed and noting that equation (6) may be written as

$$\rho = (\rho_o H_o / H)\xi^2 \tag{22}$$

to derive the following from equation (3):

$$\frac{\xi}{4H}\frac{\partial}{\partial \xi}\left(\frac{K}{H}\,\xi\,\frac{\partial T}{\partial \xi}\right) + Q_s + Q_r + Q_g = \rho\, c_p\,\frac{\partial T}{\partial t} \quad . \tag{23}$$

Note that the flow term vanishes in the constant-pressure frame. This means that no flow occurs across constant-pressure surfaces and that the atmospheric mass above a ξ-surface is constant in time. Thus, the "breathing velocity" W is automatically taken into account in this frame. An even more useful feature of this frame is the fact that one may make the approximation that the quantities (K/H) and c_p do not depend upon temperature or altitude. Replacing these by suitable averages, we rearrange the terms in equation (23) to arrive at the basic equation

$$\frac{\partial^2 T}{\partial \xi^2} + \frac{1}{\xi}\frac{\partial T}{\partial \xi} + \frac{4H}{(\overline{K/H})\xi^2}\,(Q_s + Q_r + Q_g) = \frac{1}{D}\frac{\partial T}{\partial t} \quad , \tag{24}$$

where

$$D = (\overline{K/H})g/4p_o\bar{c}_p \quad . \tag{25}$$

The heat balance equation is now uncoupled from the hydrostatic equation. Equation (24) is a partial differential equation of a well-known form in mathematical physics. Its solution permits the specification of the temperature $T(\xi,t)$ and, hence, the quantities $\rho(\xi,t)$, $z(\xi,t)$, and $W(\xi,t)$. Unfortunately, the reverse transformation in returning to the z-frame is awkward in practice, although simple in principle.

We may write the hydrostatic equation [equation (6)] in the ξ-description as

$$\frac{d\xi}{\xi} = -\frac{dz}{2H} \quad , \tag{26}$$

which upon integration gives

$$z = z_o + \int_{\xi}^{1} 2\,\frac{d\xi}{\xi}\,H(\xi,t) \quad . \tag{27}$$

The knowledge of $H(\xi,t)$ follows directly from specifying $T(\xi,t)$ and the mean molecular mass $M(\xi)$. The latter is not a function of time and is

calculated from equation (8) from the partial pressures $p_i(\xi)$. These may be calculated from a relationship between p and p_i:

$$p(\xi) = \sum_i p_i(\xi_o) \left[p_j(\xi)/p_j(\xi_o) \right]^{M_i/M_j} \quad (p_j \neq 0) \quad , \tag{28}$$

which is obvious from a comparison of equations (5), (7), and (9). Once this equation is solved (by numerical means) for the p_i's, the quantity $M(\xi)$ is determined, and hence the integration of equation (27) may be easily accomplished for each time t.

We note that the various heat sources may be easily described in the ξ-frame according to

$$Q_c = -\left(\frac{K}{H}\right)\left(\frac{\xi^2}{4H}\right)\left(\frac{\partial^2 T}{\partial \xi^2} + \frac{1}{\xi}\frac{\partial T}{\partial \xi}\right) \quad , \tag{29}$$

$$Q_s = \int d\lambda\, \eta_\lambda F_\lambda \tau_\lambda \left(\frac{\xi^2}{H}\right) \exp\left[-\tau_\lambda \xi^2 \operatorname{Ch}\chi\right] \quad , \tag{30}$$

$$Q_g^a = \left(\frac{q\xi^2}{2H}\right) \delta(\xi - \xi_m)\, f(t) \quad , \tag{31}$$

$$Q_g^b = \left(\frac{q\xi^2}{H\xi_m^2}\right) \exp\left(\frac{-\xi^2}{\xi_m^2}\right) f(t) \quad , \tag{32}$$

and

$$Q_g^c = \left(\frac{4q\xi^4}{H\xi_m^4}\right) \exp\left(\frac{-2\xi^2}{\xi_m^2}\right) f(t) \quad , \tag{33}$$

where τ_λ is the vertical optical depth above the height z_o of the atmosphere to the radiation of wavelength λ, $\tau_\lambda = \sigma_\lambda \rho_o H_o$, q is the total column rate of heat production, and ξ_m denotes the value of ξ where Q maximizes. The three types of heating sources, denoted by superscripts a, b, and c, for the geomagnetic heating are used to test the sensitivity of the solution to the shape of the height distribution.

D. Analytic Solutions for Temperature

1. STATIC SOLUTIONS

If the heat source is constant over a long time period or if we are concerned with the mean temperature of the atmosphere, we may write equation (24) as

$$\frac{\partial^2 T}{\partial \xi^2} + \frac{1}{\xi}\frac{\partial T}{\partial \xi} + \frac{4H}{(\overline{K/H})}\frac{Q}{\xi^2} = 0 \quad . \tag{34}$$

Here, Q denotes all the physical sources and sinks of energy. We may solve this equation generally in terms of Green's function $G(\xi, \xi^1)$ in which

$$\frac{\partial^2 G}{\partial \xi^2} + \frac{1}{\xi}\frac{\partial G}{\partial \xi} + \delta(\xi, \xi') = 0 \quad , \tag{35}$$

and certain boundary conditions are specified. We assume zero heat flux through the upper boundary ($\xi\, \partial T/\partial \xi \to 0$ as $\xi \to 0$) and a constant temperature T_o at the lower boundary z_o. Once G is known, the temperature at ξ is calculated for an arbitrary heating $Q(\xi)$ from

$$T(\xi) = \int_0^1 d\xi'\, \xi' \frac{4HQ(\xi')}{(\overline{K/H})\xi'^2} G(\xi, \xi') + T_o \quad . \tag{36}$$

The solution to equation (35) with the specified boundary conditions is

$$\begin{aligned} G(\xi, \xi') &= \log_e (1/\xi') \qquad \xi \le \xi' \\ &= \log_e (1/\xi) \qquad \xi \ge \xi' \quad . \end{aligned} \tag{37}$$

We display below the explicit solutions for the three heating distributions (a, b, and c), with $f(t) = 1$:

$$\begin{aligned} T^a(\xi) &= T_o + \log_e (1/\xi) \qquad \xi \le \xi_m \\ &= T_o + \log_e (1/\xi) \qquad \xi \ge \xi_m \quad , \end{aligned} \tag{38}$$

$$T^b(\xi) = T_o + \frac{q}{(\overline{K/H})}\left[\log_e (1/\xi^2) - E_1(\xi^2/\xi_m^2) + E_1(1/\xi_m^2)\right] \tag{39a}$$

$$T^b(0) = T_o + \frac{q}{(\overline{K/H})} \left[\gamma - \log_e \xi_m^2 + E_1(1/\xi_m^2) \right] \tag{39b}$$

$$T^c(\xi) = T_o + \frac{q}{(\overline{K/H})} \left[\log_e(1/\xi^2) - E_1(2\xi^2/\xi_m^2) + E_1(2/\xi_m^2) - e^{-2\xi^2/\xi_m^2} + e^{-2/\xi_m^2} \right] \tag{40a}$$

$$T^c(0) = T_o + \frac{q}{(\overline{K/H})} \left[\gamma - \log_e(\xi_m^2/2) + E_1(2/\xi_m^2) - 1 + e^{-2/\xi_m^2} \right] \tag{40b}$$

$$\gamma = 0.577216 \ldots \quad .$$

2. TIME-DEPENDENT SOLUTIONS

A general solution to equation (24) may be written in terms of the time-dependent Green function $\Gamma(\xi, \xi'; t, t')$ which obeys the equation

$$\frac{\partial^2 \Gamma}{\partial \xi^2} + \frac{1}{\xi} \frac{\partial \Gamma}{\partial \xi} + \delta(\xi, \xi') \, \delta(t, t') = \frac{1}{D} \frac{\partial \Gamma}{\partial t} \quad , \tag{41}$$

where Γ is proportional to the temperature at (ξ, t) as a result of a unit instantaneous sheet source at (ξ', t'). This equation may be found in Reference I-7 where it describes the heat flow in an infinite homogeneous cylinder of unit radius and diffusivity D. The solution is

$$\Gamma = 2D \sum_n \exp[-\alpha_n^2 D(t-t')] \, J_o(\alpha_n \xi) J_o(\alpha_n \xi') H(t-t') / J_1^2(\alpha_n) \quad , \tag{42a}$$

where $H(t-t')$ is the step function equal to 0 or 1 when $t < t'$ and $t > t'$, respectively; J_o and J_1 are Bessel functions of order 0 and 1 of the first kind; and α_n is the nth root of the Bessel function J_o. The summation is made over all roots.

Another solution to equation (41) is obtained for the case of an infinite atmosphere ($\xi_o \to \infty$); that is, for an atmosphere increasing indefinitely in mass with decreasing z. Here we impose $\xi \partial T / \partial \xi \to 0$ as $\xi \to \infty$, and the solution is

$$\Gamma = \left[\frac{1}{2} (t-t') \right] \exp \left[- \frac{\xi^2 + \xi'^2}{4D(t-t')} \right] I_o \left[\frac{\partial \xi \xi'}{4D(t-t')} \right] H(t-t') \quad , \tag{42b}$$

where I_o is a Bessel function of order zero of the first kind of imaginary argument. For small time intervals ($t \gtrapprox t'$), equation (42b) may be used to approximate equation (42a) where a large number of terms must be evaluated to obtain convergence.

A third class of solutions occurs if the lower boundary temperature is oscillatory. If ΔT is the amplitude of the temperature oscillation and if the lower boundary temperature is given by

$$T_{LB}(z_o) = T_o + \Delta T \cos[\omega(t + \phi)] \quad , \tag{43}$$

ω is the angular frequency and ϕ is the phase lag. The additional term that must be added to the general (steady-state) solution is

$$T_{LB}(\xi,t) = \Delta T \left\{ B(\xi) \cos\left[\omega(t+\phi)\right] + C(\xi) \sin\left[\omega(t+\phi)\right] \right\} \quad , \tag{44}$$

where

$$B(\xi) = \left[\mathrm{ber}(\beta\xi)\mathrm{ber}\,\beta + \mathrm{bei}(\beta\xi)\mathrm{bei}\,\beta \right] / M^2 \quad , \tag{45}$$

$$C(\xi) = \left[\mathrm{ber}(\beta\xi)\mathrm{bei}\,\beta - \mathrm{bei}(\beta\xi)\mathrm{ber}\,\beta \right] / M^2 \quad , \tag{46}$$

$$M^2 = \mathrm{ber}^2\beta + \mathrm{bei}^2\beta \quad , \tag{47}$$

$$\beta = (\omega/D)^{1/2} \quad , \tag{48}$$

where ber and bei denote the Kelvin functions [I-7]. The general solution is given by

$$T(\xi,t) = T_{LB} + \int_0^t dt' \int_0^{\xi_u} \xi' \, d\xi \; \Gamma \; 4HQ/(\overline{K/H})\xi'^2 \quad , \tag{49}$$

where ξ_u is 1 or ∞ , depending upon whether solutions (42a) or (42b) are desired.

The solutions for the time-dependent heat sources given by equations (31), (32), and (33) and for the time-dependence given by equation (17) are displayed below, where $T_{LB} = 0$:

$$T^a(\xi,t) = q\,\frac{4D}{(\overline{K/H})}\sum_n \frac{J_o(\alpha_n\xi)\,J_o(\alpha_n\xi_m)}{J_1^2(\alpha_n)}\,W_n(t,\tau) \quad , \tag{50}$$

$$T^b(\xi,t) = q\,\frac{4D}{(\overline{K/H})}\sum_n \frac{J_o(\alpha_n\xi)}{J_1^2(\alpha_n)}\,\exp\left(-\alpha_n^2\,\xi_m^2/4\right)\,W_n(t,\tau) \quad , \tag{51}$$

and

$$T^c(\xi,\tau) = q\,\frac{4D}{(\overline{K/H})}\sum_n \frac{J_o(\alpha_n\xi)}{J_1^2(\alpha_n)}\,(1-\alpha_n^2\,\xi_m^2/8)\,W_n(t,\tau) \quad , \tag{52}$$

where

$$W_n(t,\tau) = \frac{(t/\tau)\exp(-t/\tau)}{\alpha_n^2\,D\tau - 1} + \frac{\exp(-\alpha_n^2\,Dt) - \exp(-t/\tau)}{[\alpha_n^2\,D\tau - 1]^2} \quad . \tag{53}$$

Solutions for the infinite medium require a numerical integration of equation (49) except in the special case of instantaneous heating ($\tau \to 0$) where we obtain

$$T^a(\xi,t) = \frac{q}{(\overline{K/H})t}\,\exp\left[-\frac{\xi^2+\xi_m^2}{4Dt}\right]\,I_o\left(\frac{\partial\xi\xi'}{4Dt}\right) \quad . \tag{54}$$

In equations (50), (51), (52), and (54) use has been made of the definite integral

$$\int_0^\infty (\xi')^{\nu+1}\exp(-\alpha\xi'^2)\,J_\nu(\beta\xi')d\xi' = \frac{\beta^\nu}{(2\alpha)^{\nu+1}}\,\exp\left(-\frac{\beta^2}{4\alpha}\right) \quad . \tag{55}$$

For solutions (51) and (52) it is assumed that most of the heating lies above the lower boundary, so that in equation (49) we allow $\xi_u \to \infty$.

The radiative cooling term should be added to all these solutions. However, as discussed in Reference I-5, this term is of very little importance in studies of impulsive heating.

E. Application to Impulsive Heating of the Upper Atmosphere

We suppose that the atmosphere is in some quasi-steady state under the influence of the quiet-time sources and sinks of energy, which is suddenly subjected to a heat source following one of the distributions described by equations (31), (32), or (33). We have used the CIRA [I-4] model atmosphere tables to describe this quiet-time variation. We assume that the various temperature solutions describe the perturbed temperature, which amounts to invoking the linearity of the heat conduction equation. The method is to vary the parameters q, τ, and ξ_m in the postulated heat sources to examine the constraints placed upon these quantities by the satellite drag data. This analysis has been described in detail in Reference I-5, and we will review briefly those results.

Figure I-1 shows the temperature response of the atmosphere at two pressure levels. This figure illustrates the dependence of the relaxation of the temperature profile following heating by a source of duration τ of about 4 hours. The effect of the lower boundary is negligible for times less than about 1 day. Figure I-2 shows the increase in temperature at fixed altitudes as a function of time from a heat source maximizing at 150 km at a fixed local time, which is appropriate for comparison with satellite drag measurements. Figure I-3 shows the density perturbation for the same heat source as Figure I-2 at various altitudes. Quiet-time density is denoted by ρ_n.

Figure I-4 shows the temperature profile at different intervals following an instantaneous heat source, and illustrates the effect of diffusion of the injected heat from its original position and the varying speed of this diffusion as a function of height (pressure). The time-lag of an atmospheric property is defined as the interval between the time of maximization of the heat source and the time of maximization of the property. The time lags for density and temperature are shown as a function of altitude in Figure I-5. The differences between the two time lags are accounted for in Reference I-5 by the fact that the density perturbation is a combination of two processes, each of which varies in importance in both time and space.

Figure I-6 illustrates the dependence of density time lag on the heating parameter τ. For a duration of heating of about 2 hours, the time lag is of the order of 6 hours and is independent of z down to altitudes of 200 km. Figure I-7 shows the three extreme models that satisfy the satellite drag results (6 hours ± 2 hours) for the time lag following the maximum in the geomagnetic disturbance index a_p. This figure shows that if this description is valid, the heat source must be located between 140 and 160 km and must have a duration of between 1.5 and 2 hours.

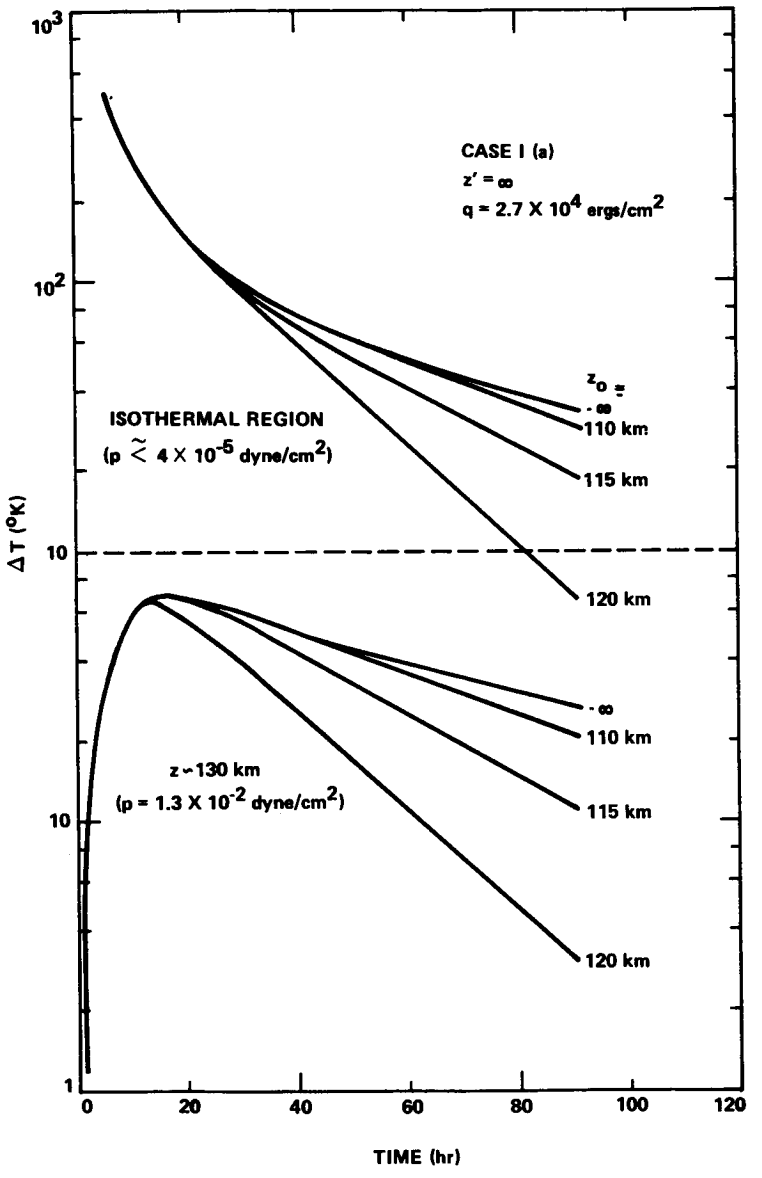


Figure I-1. Temperature variation on two constant-pressure surfaces for differing altitudes of the fixed-temperature boundary (designations (a), (b), and (c) appearing in this and all following figures refer to the particular heating distribution defined by equations (31), (32), and (33), respectively; case numbers I and II refer to the semi-infinite and infinite atmospheres, respectively).

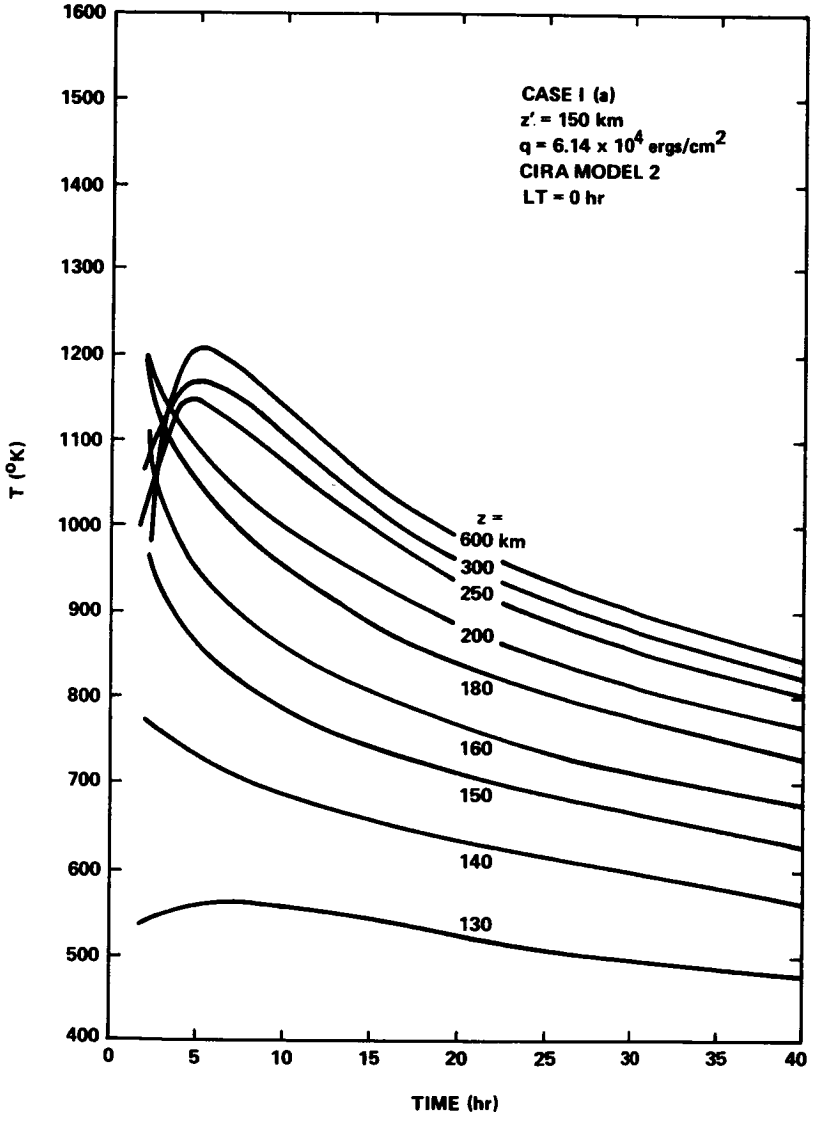


Figure I-2. Temperatures of a disturbed CIRA (Model 2) atmosphere after heating from an instantaneous sheet source initially located at 150 km (temperatures are not shown earlier than 2 hours, since they are unrealistically large in the vicinity of the heating level).

The energy requirements for a storm of maximum a_p index may be deduced from the empirical results of Jacchia et al. [1-8] for the exospheric temperature change ΔT_∞, given

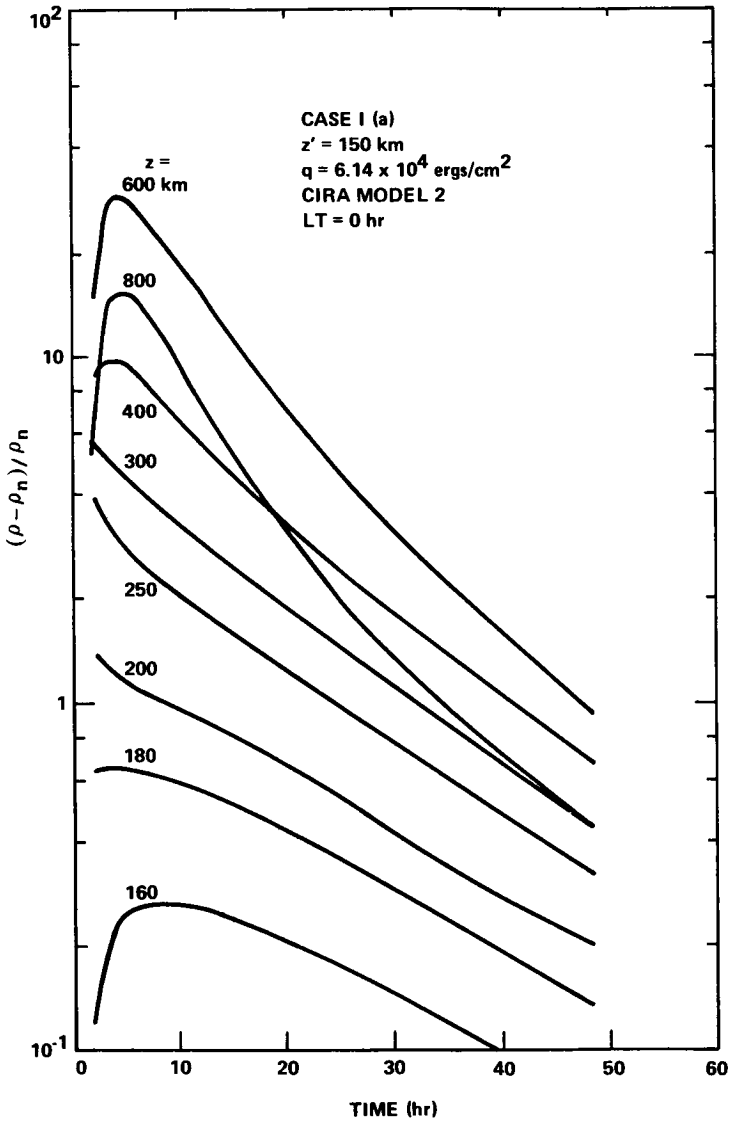


Figure I-3. Density perturbation corresponding to the temperature perturbation shown in Figure 2.

$$\Delta T_{\infty} = 1.0^{\circ}\ a_p + 100 * \left[1 - \exp(-0.08\ a_p)\right]\ ^{\circ}K \quad . \tag{56}$$

Chiu and Ching [I-9] recently pointed out that this relationship may also be accurately expressed by

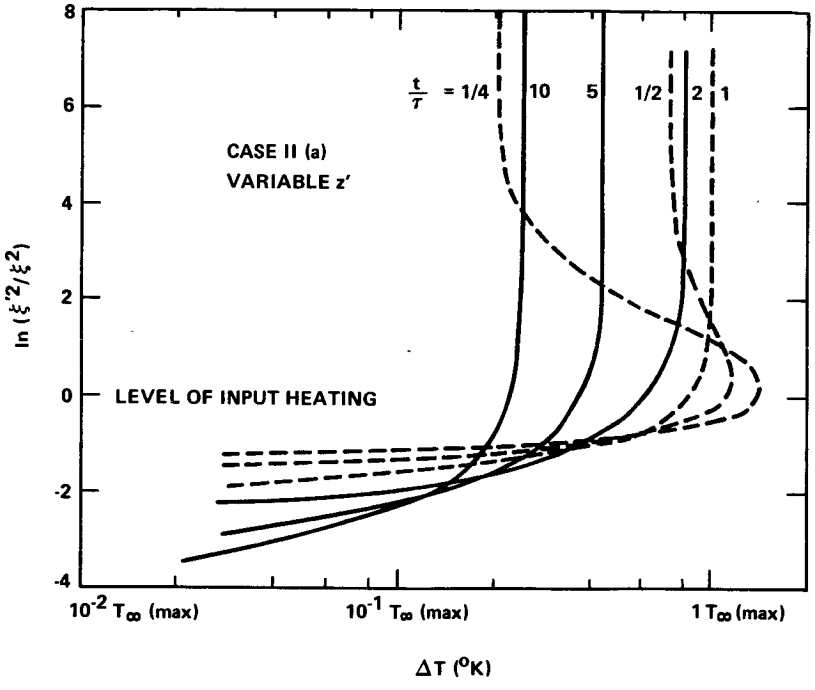


Figure I-4. Generalized temperature profiles for an instantaneous sheet source located at the level ξ', where the pressure is $P_o \xi'^2$ (τ_∞ is the time at which T_∞ maximizes, $\tau_\infty = \xi'^2/4D$; the maximum exospheric temperature is given by $T_\infty(\max) = \left[q/(\overline{K/H})\right] e^{-1}/\tau_\infty$).

$$\Delta T_\infty = (20.4)\, a_p^{0.51}\ K \quad . \tag{57}$$

The relationship between the column heating rate q and ΔT_∞ may then be used to describe the global heating rate for a given form characterized by a_p; that is,

$$q = 2600\ (a_p)^{0.51}\ \text{ergs/cm}^2 \quad , \tag{58}$$

where a_p is given in units of 2×10^{-5} gauss. This equation implies a worldwide heat input of $1.3 \times 10^{22}\ (a_p)^{0.51}$ ergs, which is an extraordinarily large number, as noted in Reference I-5.

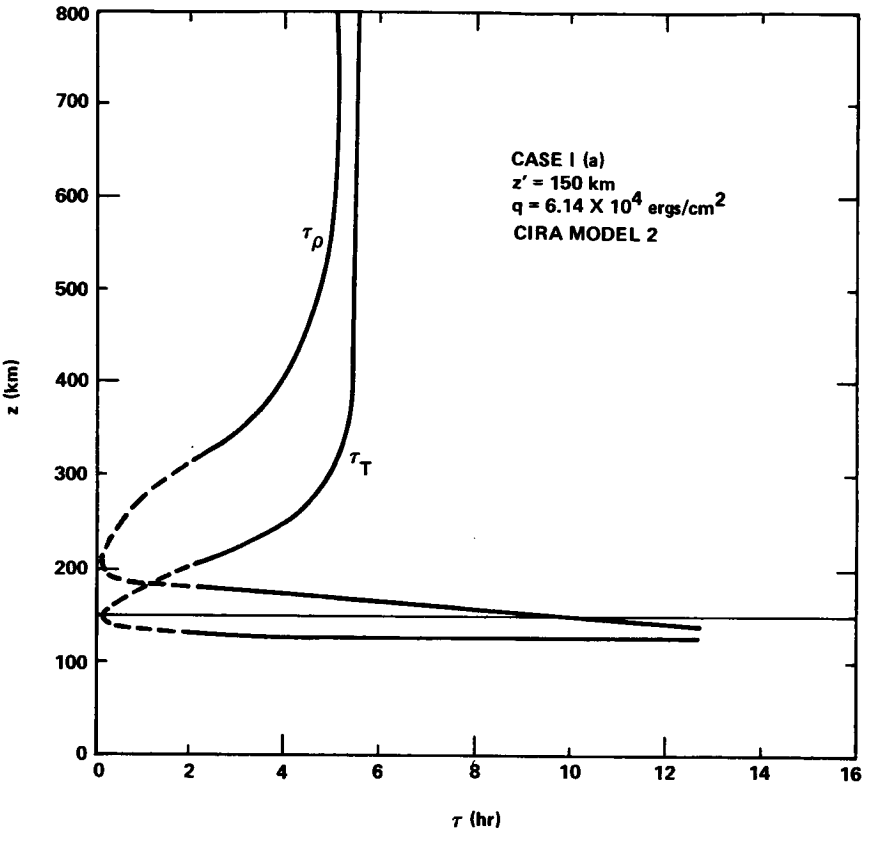


Figure I-5. Time lags of temperature and density for the temperature and density variations of Figures 2 and 3.

F. Summary and Conclusions

We have shown that analytic solutions for the temperature and density exist for a large class of problems involving impulsive heating. These methods may also be applied to the problem of the steady-state diurnal variation of temperature [I-6] and to the problem of an atmosphere whose lower boundary is undergoing temperature oscillations [I-10]. The method has also been used to examine the effects on the atmosphere of a large solar X-ray flare. Unpublished results show that, even for the largest flares that have been observed, the exospheric temperature change is no larger than about 50° K. This is a result of the deep penetration of X-rays into the high density portions of the upper atmosphere, where the specific heat per unit volume is large and temperature changes are consequently small.

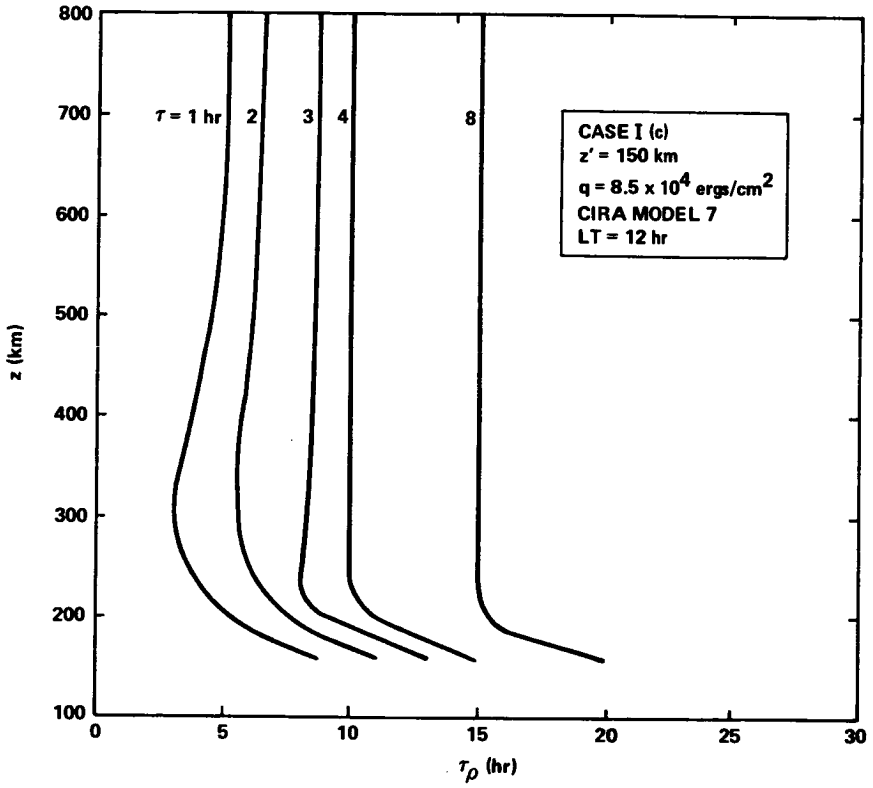


Figure I-6. Effect of the time parameter τ (half-width of the heating is approximately 1.25τ) on the time lag of density

Recently, Chiu and Ching [I-9] proposed a modification to the one-dimensional flow equation, equation (24), to allow for energy exchange of atmospheric gravity waves with the heated atmosphere. The energy exchange may be parameterized as a time-dependent diffusivity factor, whose effect is to add an additional term to equation (24). They have computed the effect on the transient response of the heated atmosphere of this additional term and have shown that significant differences appear in the final conclusions as to the spatial and temporal constraints on the geomagnetic heat source. They have shown that, whereas the energy requirements are about the same as those determined by Thomas and Ching in Reference I-5, (1) the relaxation time of the atmosphere is considerably less than that found by Thomas and Ching (Fig. I-2) and (2) the rising phase of the density response is essentially independent of the spatial character of the heating. Their results show that mechanical modes of heat loss cannot be ignored; however, their description of the cooling processes, while more realistic, is still inexact. It remains for more highly sophisticated treatments to define the properties of the

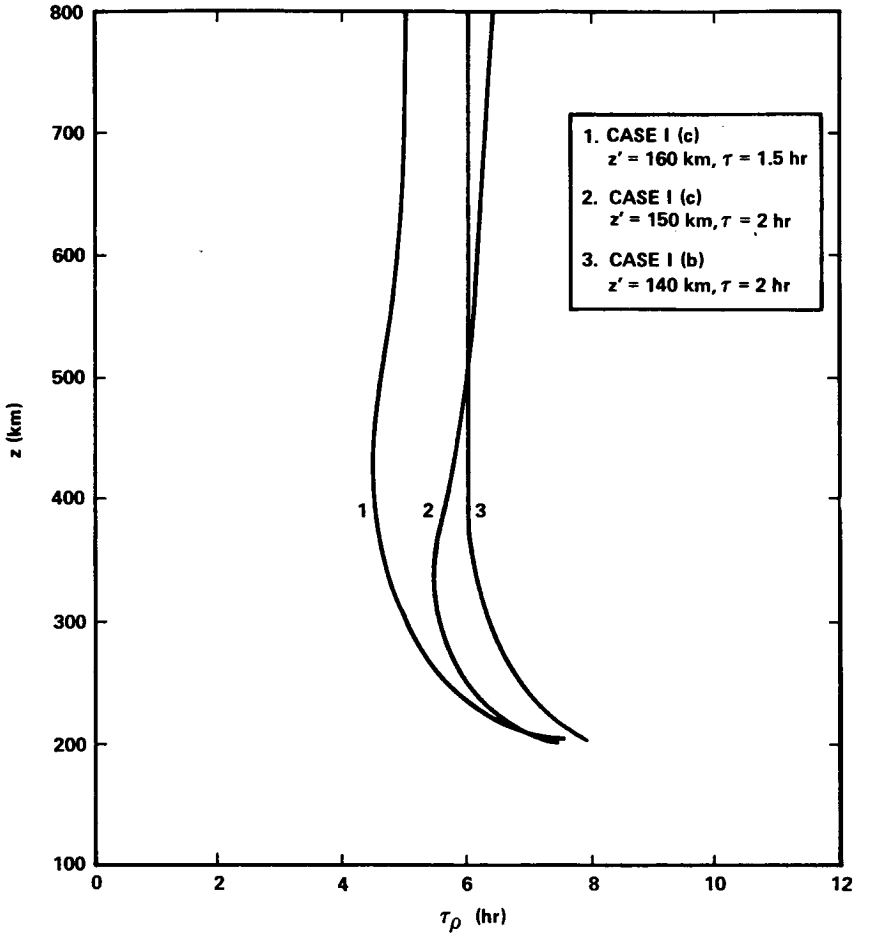


Figure I-7. Density time lags that fit satellite drag-deduced requirements (initial height assignments z' are based on CIRA Model 7, LT = 12 hours).

geomagnetic heat source. Hopefully, the present one-dimensional analytic studies will be useful in the more exact, but less intuitive, approaches.

G. References

I-1. Spitzer, L. J.: The Terrestial Atmosphere above 300 km. In The Atmospheres of the Earth and Planets, Edited by G. Kuiper, University of Chicago Press, Chicago, 1949.

I-2. Bates, D. R.: The Temperature of the Upper Atmosphere. Proc. Phys. Soc., vol. B 64, 1951, p. 805.

I-3. Harris, I.; and Priester, W.: Time dependent structure of the upper atmosphere. J. Atm. Sci., vol. 19, 1962, p. 286.

I-4. CIRA, COSPAR International Reference Atmosphere, 1965. North-Holland Publishing Company, Amsterdam, 1965.

I-5. Thomas, G. E.; and Ching, B. K.: Upper atmospheric response to transient heating. J. Geophys. Res., vol. 74, 1969, p. 1796.

I-6. Thomas, G. E.: Analytic Solutions of the Heat Conduction in the Thermosphere. Report No. TR-0158 (3260-10)-3, Aerospace Corporation, El Segundo, Calif., 1967.

I-7. Carslaw, H. S.; and Jaeger, J. C.: Conduction of Heat in Solids. Oxford, London, 1959.

I-8. Jacchia, L. G.; Slowey, J.; and Verniani, F.: Geomagnetic perturbations and upper atmosphere heating. J. Geophys. Res., vol. 72, 1967, p. 1423.

I-9. Chiu, Y. T.; and Ching, B. K.: Influence of Gravity Waves on Transient Heating Response of the Upper Atmosphere. Report No. TR-0059 (6260-10)-3, Aerospace Corporation, El Segundo, Calif., 1970.

I-10. Thomas, G. E.: The phase lags of density and temperature in the upper atmosphere. Trans. A.G.U., vol. 52, 1971, p. 481.

CHAPTER II. DISSIPATION OF ELECTRIC FIELDS IN THE IONOSPHERE

By

K. D. Cole
Division of Space Physics
La Trobe University, Bundoora, Victoria, Australia

A. Abstract

The heating and movement of the upper atmosphere at ionospheric levels in response to electric currents are discussed. Joule dissipation, generation of winds, and pressure gradients are significant factors in the energetics of the ionospheric electric currents flowing during magnetic storms and also of the Sq current system.

B. Introduction

The dissipation of the energy in steady or almost steady electromagnetic fields in the ionosphere, though intrinsically interesting, appears essential to the understanding of the energy budget of the upper atmosphere at ionospheric heights, particularly in the vicinity of the auroral zones and at times of magnetic disturbance [II-1-II-8].

This idea has had a checkered history. Chapman [II-9] refers to Schuster's early contention that the heating by upper air electric currents responsible for the solar daily magnetic variation "may help to form the isothermal layer which balloon observations have proved to exist at a height of about 50,000 feet." After the discovery of the ionosphere at considerably greater heights, where these currents must flow, Chapman [II-9] applied existing theory to the problem of joule dissipation of ionospheric electric currents and concluded that the heating attributable to Sq currents was negligible and that heating by the auroral electrojet may be about 18°K/hour. This heating is probably negligible compared to that produced by corpuscular bombardment. Modern estimates of the heating are based on Cowling's [II-10] theory of electric currents in partially ionized gas. Krassovsky [II-11] suggested anew that joule heating may be an important energy source for the thermosphere, but he did not estimate it. Dessler [II-12] considered the question partially; he concluded that joule dissipation of direct currents was not significant but that damping of hydromagnetic waves having a frequency of about 1 Hz may be.

Both the joule heating and movement caused by ionospheric electric fields have been found to be very significant to the energy budget and

dynamics of the thermosphere particularly during magnetic storms [II-1-II-3, II-5, II-7, II-8, II-13, II-14]. This conclusion is readily established using modern values of ionospheric electrical conductivity and estimates of steady electromagnetic fields in the ionosphere.

A large number of workers classify all heating of the thermosphere that is not associated with extreme ultraviolet solar radiation [II-15] as a corpuscular effect. However, there are at least three significant categories of heating involved in this blanket term; namely, (a) truly corpuscular bombardment of the atmosphere as that produced by auroral electrons and protons [II-16] and possibly neutral atoms, (b) joule dissipation of electric currents [II-1], and (c) conduction of heat from the magnetosphere [II-17]. A fourth possibly important category is dissipation of hydromagnetic waves [II-12]. Inferences of a corpuscular effect on the thermosphere have been made by a number of workers (see Reference II-18 for a review). This paper is basically concerned with joule dissipation of electric currents in the ionosphere; however, it also considers the coupled problem of movement of the thermosphere by Lorentz forces.

C. Theory

This paper discusses some effects of electric fields applied to the ionosphere. (Electromagnetic units are used.) Baker and Martyn [II-19], Fejer [II-20], and Hirono [II-21] did much to place the electrodynamics of the ionosphere on its present footing. The electric current in the ionosphere may be written [II-22] as

$$\underline{j} = \sigma_0 \underline{E'}_{//} + \sigma_1 \underline{E'}_{\perp} + \sigma_2 \frac{\underline{B} \times \underline{E'}}{B} \quad , \tag{1}$$

where

$\underline{E'} = \underline{E} + \underline{V} \times \underline{B}$,

$\underline{V}$ = the momentum (per unit mass) of ionospheric gas,

$\underline{B}$ = the geomagnetic field,

$\underline{E}$ = the electric field in the rest frame of reference,

σ_0 = the conductivity parallel to $\underline{B}$,

σ_1 = the Pedersen conductivity,

σ_2 = the Hall conductivity,

and

// and ⊥ indicate components of a vector parallel and perpendicular, respectively, to $\underline{B}$.

The theory of conductivity appropriate for application to the ionosphere has been developed by Cowling [II-10] and more recently by Burgers. Assuming the ionosphere to be representable locally by a partially ionized ternary gas from one species of neutral particles and adopting a free-path method of treatment, Baker and Martyn [II-19] showed

$$\sigma_0 = n_e e^2 \left(\frac{1}{m_e \nu_e} + \frac{1}{m_i \nu_i} \right) \tag{2}$$

$$\sigma_1 = n_e e^2 \left[\frac{\nu_e}{m_e(\omega_e^2 + \nu_e^2)} + \frac{\nu_i}{m_i(\omega_i^2 + \nu_i^2)} \right] , \tag{3}$$

and

$$\sigma_2 = n_e e^2 \left[\frac{\omega_e}{m_e(\omega_e^2 + \nu_e^2)} - \frac{\omega_i}{m_i(\omega_i^2 + \nu_i^2)} \right] , \tag{4}$$

where

n_e = the number of electrons cm^{-3},

e = the electronic change,

$\nu_{e,i}$ = the electron or ion collision frequency with neutrals,

$m_{e,i}$ = the mass of an electron or ion,

and

$\omega_{e,i} = Be/m_{e,i}$.

If the ionosphere is regarded more precisely as a partially ionized gas mixture, then the contributions to the conductivity parameters by each ion must be summed, as has been done by Rees and Walker [II-23]. Banks [II-24, II-25] has reviewed the appropriate expressions to be used for electron and ion collision frequencies when Maxwellian velocity distributions of charged particles and of neutrals may be assumed.

This paper concentrates on the physical principles of dissipation of electric fields. In the ionosphere, $\sigma_0 >> \sigma_1$ or σ_2 [II-22]. As a

consequence of this inequality, $E_{//} << E_{\perp}$, where // and $\perp$ stand for parallel and perpendicular to the geomagnetic field. For many purposes in the ionosphere and magnetosphere, it is common and indeed adequate to assume $E_{//} = 0$. In the first instance, we shall consider the situation in which $E_{//} = 0$. Consider a uniform partially ionized ternary gas that simulates the ionosphere at some altitude of interest. Piddington [II-26] has dealt with this situation and the following is his basic approach.

The equation of electric current may be written as

$$\underline{j} = \sigma_1 \underline{E}' + \sigma_2 \frac{\underline{B} \times \underline{E}'}{B} \quad , \tag{5}$$

where

$$\underline{E}' = \underline{E} + \underline{V} \times \underline{B}$$

and

$$E_{//} = 0 \quad .$$

For a fluid medium, we may write

$$\rho \frac{d\underline{V}}{dt} = \underline{F} + \underline{j} \times \underline{B} \tag{6}$$

where

$$\frac{d\underline{V}}{dt} = \frac{\partial \underline{V}}{\partial t} + (\underline{V} \cdot \nabla)\underline{V} \tag{7}$$

ρ = the mass density of the fluid,

and

$\underline{F}$ = the mechanical force per unit volume.

It is assumed that $F_{//} = 0$ and $V_{//} = 0$. From equations (5) and (6), it follows that

$$\underline{V} = \frac{\underline{E} \times \underline{B}}{B^2} + \frac{R_1 \rho}{B^2}\left(\underline{F} - \frac{d\underline{V}}{dt}\right) + R_2 \frac{\underline{B} \times \left(\underline{F} - \rho \frac{d\underline{V}}{dt}\right)}{B^3} \quad , \tag{8}$$

where

$$R_1 = \frac{\sigma_1}{\sigma_1^2 + \sigma_2^2} \tag{9}$$

and

$$R_2 = \frac{\sigma_2}{\sigma_1^2 + \sigma_2^2} \quad . \tag{10}$$

The instantaneous work done per unit volume per second against the force $\underline{F}$ is given by

$$-\underline{F} \cdot \underline{V} = -\underline{F} \cdot \frac{\underline{E} \times \underline{B}}{B^2} - \frac{R_1 F^2}{B^2} + \frac{R_1 \rho}{B^2}\ \underline{F} \cdot \frac{d\underline{V}}{dt} + \frac{R_2}{B^3}\ \underline{F} \cdot \underline{B} \times \rho \frac{d\underline{V}}{dt} \quad . \tag{11}$$

The heat produced per unit volume per second by joule dissipation of electric current is given by

$$\underline{j} \cdot \underline{E}' = \frac{R_1}{B^2}\left(F - \rho \frac{d\underline{V}}{dt}\right)^2 = \frac{R_1}{B^2}\ F^2 + 2\frac{R_1 \rho^2}{B^2}\left(\frac{d\underline{V}}{dt}\right)^2 - 2\frac{R_1 \rho}{B^2}\ \underline{F} \cdot \frac{d\underline{V}}{dt} \quad . \tag{12}$$

Alternatively [II-1],

$$\underline{j} \cdot \underline{E}' = \sigma_1 E'^2 = R_1 j^2 = \frac{j^2}{\sigma^3} \quad , \tag{13}$$

where

$\sigma_3 = \sigma_1 + \sigma_2^2/\sigma_1$ (the Cowling conductivity).

The total work done per unit volume sec^{-1} by the moving conducting medium is given by

$$W = -\underline{F} \cdot \underline{V} + \underline{j} \cdot \underline{E}' \quad . \tag{14}$$

When F and E are uniform and E is applied abruptly at some instant, Piddington [II-26] showed that the medium responds like a damped harmonic oscillator. The period of the oscillations is

$$T_P = 2\pi\rho/\sigma_2 B^2 \quad , \tag{15}$$

and the characteristic damping of the oscillation is

$$T_D = \rho/\sigma_1 B^2 \quad . \tag{16}$$

Values of T_P and T_D at various altitudes in a model ionosphere are shown in Table II-1. The table also shows values of σ_0, σ_1, and σ_2 [II-22].

TABLE II-1. ATMOSPHERIC PARAMETERS

	Altitude (km)				
	100	125	150	175	200
Electron density (cm^{-3})	1(5)[a]	1.5(5)	2(5)	2(5)	2(5)
T_P(sec)	6(6)	2.7(5)	1.3(5)	4(5)	6(5)
T_D(sec)	2.7(7)	1.5(5)	1.2(4)	1(4)	8(3)
σ_0	1.7(-13)	2.9(-12)	1.6(-11)	4.1(-11)	7.7(-11)
σ_1	1.2(-16)	1.3(-15)	2.9(-15)	1.1(-15)	3.9(-16)
σ_2	3.3(-15)	4.5(-15)	1.7(-15)	1.7(-16)	2.3(-17)

a. Numbers in parenthesis indicate powers of 10 by which the numbers outside the parenthesis are to be multiplied.

The quantities T_P and T_D are inversely proportional to the assumed electron density. These parameters give an estimate of likely response times of the ionosphere to an impulsively applied electric field, such as those occurring during some magnetic disturbance [II-27]. The ionosphere may be said to be generally either overdamped or critically damped. T_D indicates that, in the case of electric fields being applied to the ionosphere for times up to 3 hours, the atmosphere above 150 km may be moved at the final steady-state speed. Below this altitude, a steady state could not be reached by fields of such duration. In this event the joule dissipation is best calculated by equation (13) because there are magnetic observations that can be used to infer j in auroral electrojets [II-1, II-28 - II-30]; whereas, a lack of relevant observations makes equation (12) not directly useful except perhaps in numerical model calculations. Even though a steady state may not be reached in the lower ionosphere, such a state may be attained in the upper ionosphere. In the particular circumstances of a steady state and either $|\rho(\underline{V} \cdot \nabla)\underline{V}| << |F|$ or the fluid being incompressible (refer to Reference II-26),

$$\underline{V} = \frac{\underline{E} \times \underline{B}}{B^2} + \frac{R_1 \underline{F}}{B^2} + R_2 \frac{\underline{B} \times \underline{F}}{B^2} \quad ; \tag{17}$$

and, in this case,

$$W = -\underline{F} \cdot \underline{E} \times \underline{B}/B^2 \tag{18}$$

and

$$\underline{j} = \frac{\underline{F} \times \underline{B}}{B^2} \quad . \tag{19}$$

During a magnetic disturbance, both electric fields of external origin [II-7, II-27, II-31] and local dynamo fields [II-7, II-27, II-32] may drive electric currents in the ionosphere. As a result of heating of the atmosphere, gradients of pressure may also play a role at some stage of the disturbance [II-1]. In what follows, it is assumed that only fields of external origin are applied and these are such that $E_{//} = 0$, or approximately so, and one considers the effects of E which is applied for a time t_E. The electric field E is supposedly maintained by current flow along the geomagnetic field lines from a magnetospheric source (refer to Reference II-27), and it finds continuity with the Pedersen current in the ionosphere. In addition, Hall currents flow, and these cause most of the magnetic disturbance observed at the earth's surface [II-33], after due allowance for currents induced in the ground and sea. Two important cases will be considered: (a) $t_E \ll T_D$ and (b) $t_E \gg T_D$.

D. Joule Heating — Case a: $t_E \ll T_D$

This case applies to the lower ionosphere; i.e., altitudes up to about 140 km. Imagine an auroral ionosphere consisting of a series of parallel plane laminas of auroral ionization. The surfaces of the laminas are parallel to $\underline{B}$. Current flows along the geomagnetic field outside these slabs of auroral ionization, and Pedersen current flows across the slabs perpendicular to their surfaces. The geomagnetic field lines are regarded as approximately equipotentials of any electrostatic field because $\sigma_0 \gg \sigma_1$. The source current from the outer magnetosphere is assumed to flow to the ionosphere in a sheet parallel to and outside of the slabs of ionization. The return current to the magnetosphere flows up the field lines in a sheet parallel to the source current so that the auroral slabs are contained between the source and return current sheets. The Pedersen current flow orthogonal to the auroral slabs is then given by

$$\sum_1 E = J_P = \text{constant} \quad , \tag{20}$$

where

$$\sum_1 = \int \sigma_1 \, d\ell \quad .$$

Here the integral is taken along the geomagnetic field line through the ionosphere:

$$J_P = \int \sigma_1 E \, d\ell \quad .$$

It follows that

$$E = \frac{\text{constant}}{\sum_1} \quad , \tag{21}$$

and this equation conforms to the observation [II-34, II-35] that electric fields are stronger outside some auroral forms than within them. The volume joule heating rate Q(J) is best calculated by j^2/σ_3 or equivalently $\sigma_1 E^2$. Thus,

$$Q(J) = \sigma_1 E^2 = \sigma_1 \frac{J_P^2}{\sum_1^2} \quad , \tag{22}$$

so that Q(J) would be distributed in altitude the same way as σ_1. The integrated heating rate is given by

$$\int Q(J) \, d\ell = \frac{J_P^2}{\sum_1} \tag{23}$$

$$= \sum_1 E^2 \quad .$$

On this model, the heating outside of the auroral slabs is greater than that inside. This conclusion, however, may not be general. In a sunlit auroral ionosphere, the conductivity may be fairly uniform. Adopting a quiet-day value of σ_1 of 10^{-8} emu [II-22], the following daytime integrated joule heating rates would apply (Table II-2). These numbers may be compared to the heat requirement of the quiet atmosphere at these altitudes, which is in the range of 1 to 10 ergs cm^{-2} sec^{-1} [II-36].

TABLE II-2. JOULE HEAT

E(mV/m)	10	30	50	100
$\int Q(J)\,d\ell$ ergs cm^{-2} sec^{-1}	1	10	25	100

At night the ionosphere between auroral forms is probably considerably less dense than within auroras so that for the model of the auroral ionosphere considered, the joule heating will be much more intense outside auroral forms than within them.

From equation (22) the column heating rate is given by

$$\left(\sum_{1_a} \sum_{E_a}\right)^2 \Big/ \sum_1 \quad ,$$

where, for convenience, one may take $\sum_{1_a}$ and E_a to be referring to inside an auroral form, the product of those two quantities being constant. It is convenient to have an estimate of the joule heating in terms of magnetic disturbance ΔH as measured at the ground.

We can do this as follows. First,

$$J(HALL) = \frac{\sum_2}{\sum_1} K \quad , \tag{24}$$

where

$$K = \sum_1 E = J_P = \text{constant.}$$

Let us make the assumption that $\sum_2 \Big/ \sum_1$ is constant inside and outside auroral forms. This is probably a crude but useful assumption. Then, J(Hall) equals the constant inside and outside of auroral forms in which case, for a broad auroral zone,

$$\Delta H = (2/3)\ (4\pi)\ \sum_2 E \quad , \tag{25}$$

where the fraction 2/3 takes into account the induction in the earth. Equation (23) then yields

$$\int Q(J)\,d\ell = \frac{g}{(8\pi)^2} \frac{\sum_1 (\Delta H)^2}{\sum_2^2} \text{ ergs cm}^{-2} \text{ sec}^{-1} \quad . \tag{26}$$

This formula should give a rough guide to the magnitude of joule heating by auroral electrojets. Assuming $\sum_1 \sum_2 \approx \frac{1}{2}$ [II-22] and that outside auroral forms $\sum_2 = \frac{1}{10}$ that of daytime; i.e., $\sum_2 \approx 10^{-9}$ emu,

$$\int Q(J)\,d\ell = 7 \times 10^{-4} (\Delta H)^2 \text{ ergs cm}^{-2} \text{ sec}^{-1} \quad , \qquad (27)$$

where ΔH is in gammas. This is comparable to the estimate made earlier; namely, Model B of Reference II-7.

Formula 27 gives rise to Table II-3.

TABLE II-3. JOULE HEAT AND WIND SPEED

ΔH (gamma)	40	100	500	1000	2000
$\int Q(J)\,d\ell$ ergs cm^{-2} sec^{-1}	1	7	175	700	2800
Wind Speed (m sec^{-1})	4.2	11	55	110	440

The energy deposited in the lower ionosphere by joule heating will be dissipated by generation of winds, lifting of the atmosphere, and heat conduction. It is probable that the air will rise over the heated region and rush in from the sides to form giant convection cells on either side. Suppose the latitudinal dimension of the convection cells is L_1 and the latitudinal dimension of the auroral electrojet is L_2. If one-half of the joule heating for 1 hour were to go into generating such convection cells above, for example, 100 km altitude, the typical wind speed would be given by

$$v^2 = \frac{1800\, L_2}{L_1 M} \int Q(\ell)\, d\ell \quad ,$$

where

M = the mass of gas cm^{-2} above 100 km altitude.

The values of v in m/sec^{-1}, assuming $L_1/L_2 = 10$ and $M = 10^{-3}$ gm, are shown in Table II-3. These values are significant when compared to normal wind speeds (approximately 30 m/sec) at these heights. Such winds would lead to important secondary electromotive forces generated in the auroral ionization, leading to additional current flow and severe modification to electric potentials in the magnetosophere.

Similar wind speeds were deduced earlier [II-1] on the assumption that a steady state existed. It is possible that intense joule heating could

shorten the time required to reach a steady state to much less than T_D. Numerical studies combining the equations of motion and energy are required. These studies are in progress.

E. Winds and Possible Pressure Gradients — Case b: $t_E \gg T_D$

This case applies to the upper ionosphere above approximately 140 km for electric fields of duration in excess of a few hours. Because σ_1 determines the time t_E, a steady state will be reached in high conductivity regions before it will be reached in regions of less conductivity. For this reason, great velocity shears would exist in the auroral atmosphere above approximately 130 km altitude. In the absence of a significant mechanical force, the final velocity of the neutral gas is $\underline{E} \times \underline{B}/B^2$. Table II-4 shows values of the steady-state wind V_W, assuming B = 0.5 gauss.

TABLE II-4. ELECTRIC FIELDS AND NEUTRAL DRIFTS

E(mV/m)	10	50	100	150
V_W(m/sec)	200	1000	2000	3000

Such winds could impart considerable momentum to the upper atmosphere during severe, long-lasting storms. It so happens that negative magnetic bays [II-2] are larger and stronger than positive magnetic bays at the auroral zone; therefore, a new westward movement of the upper atmosphere would tend to be created at auroral latitudes. The momentum associated with this movement may be transported to lower latitudes by the operation of a recently postulated eddy exchange coefficient [II-8] and contribute to the observed superrotation of the midlatitude upper atmosphere [II-37].

Because of joule heating, however, a significant gradient of pressure in the neutral gas may build up over the auroral region. In this event $\underline{E}' = \underline{E} + \underline{V} \times \underline{B} \neq 0$ above 130 km altitude, and from (17)

$$\underline{E}' = R_1 \frac{\underline{F} \times \underline{B}}{B^2} + \frac{R_2 \underline{F}}{B} \quad . \tag{28}$$

From this, it follows that

$$\underline{F} = (\sigma_1 \sigma_3)^{1/2} E' B \quad . \tag{29}$$

Values of E up to 100 mV/m (or 10^5 emu) have been measured in the auroral atmosphere [II-35]. E provides an upper limit to E' because it is likely that $\underline{V} \times \underline{B}$ subtracts from $\underline{E}$. If $\underline{F}_\perp$ derives from a gradient in pressure ($\underline{\nabla P}$) in the neutral gas, then an upper limit to ∇_p is given by equation (29) for E'. It would be advantageous to measure $\underline{E}$ and $\underline{V}$ at these altitudes to ascertain $\underline{E}'$.

Consider the equation

$$\nabla_p = (\sigma_1\sigma_3)^{1/2} E' B \qquad (30)$$

and what it would mean at an altitude of, for example, 200 km across an auroral zone 500 km wide. Assuming B = 0.5, $n_e = 5 \times 10^5$ cm^{-3}, and $(\sigma_1\sigma_3)^{1/2} = 10^{-15}$ emu [II-22], equation (3) implies pressure differences (δ_p) across the auroral zone given by Table II-5.

TABLE II-5. PRESSURE PERTURBATIONS

E' (mV/m)	10	50	100
δ_p dynes cm^{-2}	2.5×10^{-4}	1.25×10^{-3}	2.5×10^{-3}

These values of δ_p may be compared to the pressure at 200 km which, in one model atmosphere [II-22], is estimated to be approximately 2×10^{-3} dynes cm^{-2}.

F. The Sq System

Thus far this discussion has hinged around the auroral electrojet system. It is of interest to apply similar considerations to the Sq current system. By virtue of the good conductivity along the geomagnetic field, large-scale electrostatic fields generated in the E-region by dynamo action are conveyed to the F-region, where they cause the neutral gas to move. In the absence of a net mechanical force in the F-region the neutral velocity would become $\underline{E} \times \underline{B}/B^2$. However, a non-zero value of $\underline{E}' = \underline{E} + \underline{V} \times \underline{B}$ in a steady state would indicate the presence of Coriolis and pressure gradient forces. Suppose $E' \approx 10$ mV/m $\approx E$ and that this is explainable by a pressure gradient force. This would mean, from equation (3) and assuming $n_e = 5 \times 10^5$ cm^{-3}, that $\nabla_p = 5 \times 10^{-12}$ dynes cm^{-2} at an altitude of 200 km. Across the dimension of an Sq current cell of about 3000 km, there could exist a pressure difference at 200 km altitude of 1.5×10^{-3} dynes cm^{-2}. This difference is comparable to the pressure at that altitude. We reach the conclusion that, in the event that $\underline{E} + \underline{V} \times B \neq 0$, in the F-region, the difference may be accounted for in

terms of pressure differences on a global scale equal to the pressure at the height concerned. Experiments designed to measure $\underline{E}$, $\underline{V}$, and P simultaneously at widely spaced regions of the globe could test this idea.

The joule heating associated with the Sq current system has been discussed before [II-3] and found to be about 10^{-8} ergs cm^{-3} sec^{-1} in the height range. Integrated over the height range of 50 to 100 km, this heating amounts to about 0.5 ergs cm^{-2} sec^{-1} which is comparable to the heat flux conducted downward at the base of the thermosphere [II-36].

G. Conclusions

It is clear that heating and movement of the upper atmosphere at ionospheric levels by means of electric fields is of importance not only for Sq fields but also for fields established at times of magnetic disturbance. The most dramatic effects are associated with the auroral electrojet system.

The following effects are known to exist because they are so large or already have been observed.

1. Movement of the atmosphere at F-region altitudes caused by electric fields of the Sq current system [II-19, II-37, II-38].

2. Movement of the atmosphere at F-region altitudes caused by the electric fields of the auroral electrojet system [II-1, II-8].

3. Generation of significant pressure, density, wind, and temperature changes in the thermosphere [II-1-II-4, II-5, II-7] by disturbance ionospheric currents.

Effects that may be more difficult to determine by experimental means are associated, in part, with the joule heating of the atmosphere by the Sq current system. These include a component of the pressure differences across Sq current cells, thermospheric heights, and also changes in density and temperature. These effects would be superimposed on and may be of similar order to normal diurnal variations caused by solar radiation.

There are large areas available for new work, particularly in numerical modeling of the effects described. Further experimentation is required to determine directly such parameters as wind velocity, pressure, temperature, and electric field under both quiet and disturbed magnetic conditions. Much has already been learned by experimental and theoretical means [II-18, II-39 - II-43], but much more remains to be done.

H. References

II-1. Cole, K. D.: J. Phys. Soc. Japan, vol. 17, supp. A-I, part I, 1962, pp. 296-300.

II-2. Cole, K. D.: Australian J. Phys., vol. 15, 1962, pp. 223-235.

II-3. Cole, K. D.: Nature, vol. 194, 1962, p. 42.

II-4. Cole, K. D.: Nature, vol. 194, 1962, p. 75.

II-5. Cole, K. D.: J. Geophys. Res., vol. 68, 1963, pp. 3231-3235.

II-6. Cole, K. D.: Nature, vol. 199, 1963, pp. 444-445.

II-7. Cole, K. D.: Planet. Space Sci., vol. 10, 1963, pp. 129-163.

II-8. Cole, K. D.: Planet. Space Sci. vol. 19, 1971.

II-9. Chapman, S.: Terr. Mag. Atmos. Elec., vol. 42, 1937, p. 355.

II-10. Cowling, T. G.: Proc. Roy. Soc. (London), vol. 183, 1945, p. 453.

II-11. Krassovsky, V. I.: Planet. Space Sci., vol. 1, 1959, p. 14.

II-12. Dessler, A. J.: J. Geophys. Res., vol. 64, 1959, pp. 389-401.

II-13. Kato, S.: Planet. Space Sci., vol. 9, 1962, p. 939.

II-14. Rubashev, B. M.: NASA TTF-244, 1964.

II-15. Hinteregger, H. E.; Hall, L. A.; and Schmidthe, G. S.: Space Res., vol. V, 1965, p. 1175.

II-16. Dalgarno, A.: Annales de Geophysique, January-March, vol. 20, 1964, pp. 65-74.

II-17. Cole, K. D.: J. Geophys. Res., vol. 70, pp. 1689-1706.

II-18. Roemer, M.: Ann. de Geophys., vol. 25, 1969, pp. 419-437.

II-19. Baker, W. G.; and Martyn, D. F.: Phil. Trans. Roy. Soc. London, vol. A 246, 1953, p. 281.

II-20. Fejer, J. A.: J. Atmos. Terr. Phys. vol. 4, 1953, p. 184.

II-21. Hirono, M.: J. Geomag. Geoelect, Kyoto, vol. 5, 1953, p. 22.

II-22. Chapman, S.: Il. Nuovo Cimento Supplement No. 4, vol. 4, series X, 1956, pp. 1385-1412.

II-23. Rees, M. H.; and Walker, J. C. G.: Ann. de Geophys., vol. 24, 1968, pp. 1-7.

II-24. Banks, P.: Planet. Space Sci., vol. 14, 1966, pp. 1105-1122.

II-25. Banks, P.: Planet. Space Sci., vol. 14, 1966, pp. 1085-1103.

II-26. Piddington, J. A.: Mon. Notices Roy. Astron. Soc., vol. 114, pp. 651-663.

II-27. Cole, K. D.: Geophys. J. Roy. Atr. Soc., vol. 6, 1961, pp. 103-114.

II-28. Chapman, S.; and Bartels, J.: Geomagnetism. Vols. I and II, Clarendon Press, Oxford, 1962.

II-29. Armstrong, J. C.: Trans. Amer. Geophys. Un., vol. 51, 1970, p. 405.

II-30. Potter, W. E.; and Cahill, J. R.: J. Geophys. Res., vol. 74, no. 21, 1969.

II-31. Axford, W. I.; and Hines, C. O.: Can. J. Phys., vol. 39, 1961, p. 1433.

II-32. Cole, K. D.: Australian J. Phys., vol. 13, 1960, pp. 489-497.

II-33. Heppner, J. P.: Goddard Space Flight Center, X-612-68-331, Aug. 1968.

II-34. Aggson, T. L.: Atmospheric Emissions. Edited by B. M. McCormac and A. Omholt, D. Van Nostrand Company, Inc., Princeton, N.J., 1969.

II-35. Aggson, T. L.: Proc. Spring A.G.U. Meeting, Washington, D.C., 1970.

II-36. Nicolet, M.: In Physics of the Upper Atmosphere. Edited by J. A. Ratcliffe, Academic Press, New York, N.Y., 1960.

II-37. King-Hele, D. G.: Planet. Space Sci., vol. 12, 1964, pp. 835-853.

II-38. Challinor, R. A.: Planet. Space Sci., vol. 18, 1970, p. 1485.

II-39. Jacchia, L. G.: Nature, vol. 183, 1959, pp. 526 and 1662.

II-40. Jacchia, L. G.: Planet. Space Sci., vol. 12, 1964, pp. 355-378.

II-41. Jacchia, L. G.; and Slowey, J.: J. Geophys. Res., vol. 69, 1964, p. 905.

II-42. Moe, K.: Nature, vol. 201, 1963, p. 481.

II-43. De Vries, L. L.; Friday, E. W.; and Jones, L. C.: Space Research, vol. 7, 1967, p. 1173.

CHAPTER III. STRUCTURE AND VARIATIONS OF THE HETEROSPHERE[1]

By

L. G. Jacchia
Smithsonian Astrophysical Observatory,
Cambridge, Massachusetts

A. Abstract

Atmospheric composition is kept almost homogeneous to a height of approximately 90 km (the homosphere) by mixing; above 100 km (the heterosphere), diffusive separation of the atmospheric constituents occurs. The heterosphere is dominated by two energy sources: solar extreme ultraviolet radiation and the solar wind. Many large variations are observed in this region and they can all be related, directly or indirectly, to variations in the intrinsic intensity of these two sources and to variations in the location of the sources relative to the observer. Atmospheric models can be successfully constructed in which the temperature and density variations are empirically related to solar and geomagnetic parameters. Some of the more recent findings concerning the structure and variations in the heterosphere are presented.

B. Discussion

From the viewpoint of composition, the earth's atmosphere can be divided into two regions: (1) the homosphere, which extends from the earth's surface to a height of approximately 100 km and in which mixing keeps the composition practically constant, and (2) the heterosphere, above 100 km, where, because of low collision frequency, mixing is ineffective and diffusion equilibrium prevails. The homopause, which separates these two regions, is somewhat variable in height and cannot be considered as a sharp dividing surface, although observations indicate that the transition zone is relatively thin. Photodissociation of molecular oxygen occurs in a region very close to the homopause, with the result that atomic oxygen becomes an important constituent of the heterosphere; in fact, atomic oxygen is the most important constituent in the region from 200 to 600 km. In diffusion equilibrium each atmospheric constituent has its own individual scale height that is determined by its molecular or atomic mass and by the temperature; therefore, as the temperature varies,

1. This work was supported in part by NASA grant NGR 09-015-002.

the composition of the heterosphere also varies as does its total density at any given height. The density variations are largest at a height of 500 to 600 km where they reach a factor of about 300. Most of our knowledge on these variations has come from satellite-drag analyses.

The main source of energy in the heterosphere is provided by solar extreme ultraviolet (XUV) radiation, which is greatly variable in the course of an 11-year cycle of solar activity and with the formation and disintegration of individual active areas on the sun. In the heterosphere the temperature rises very rapidly in the region where this radiation is absorbed (100 to 200 km), then rises more slowly until it reaches an asymptotic value ("exospheric temperature") that is primarily determined by the level of XUV radiation but is influenced also by other factors. Since different spectral emissions are involved in the radiation from the clear disk of the sun and from active areas, we must distinguish, in the solar activity effect on the atmosphere, between a slow disk component and an erratic active-area component, the latter being affected by the 27-day synodic rotation of the sun. Quantitative data on solar XUV radiation are not continuously available, so the 10.7-cm flux from the sun, which is easily measured on the ground and whose variations closely parallel those of the XUV radiation, is generally used with excellent results to predict atmospheric temperature variations. Exospheric temperatures range from a minimum of about 550°K at sunspot minimum to somewhere in excess of 2000°K under special conditions at extremely high levels of solar activity.

As the absorption of solar XUV changes in the course of the day and stops at night, in a pattern that varies with the latitude and the seasons, we observe a season- and latitude-dependent diurnal variation that reaches a maximum of about 30 percent in the temperature and is fairly amenable to prediction through empirical models (less so by purely theoretical models).

Another source of energy in the heterosphere comes from the solar wind, whose fluctuations, as evidenced by the variations of the geomagnetic field, are faithfully reflected in the variations of temperature and density. First-class magnetic storms are accompanied by increases in density that may reach one order of magnitude at a height of 500 km, but even the smallest geomagnetic variations have their counterpart in the atmosphere. Recent high-resolution observations with satellite-borne accelerometers seem to indicate that the perturbations start in a segment of the auroral belt and are carried to the rest of the globe by traveling gravity waves. Joule currents are suspected as the prime agent of the observed atmospheric heating. More mysterious is a global semiannual density variation, which is always present although with a somewhat variable amplitude. The maxima and minima are unequal: a deep minimum in July is followed by a high maximum in October; the January minimum and the April maximum are less pronounced. The density variations do not seem

to be accompanied by important variations in the temperature and are observed at all heights above 90 km with a maximum amplitude at about 800 km.

Helium dominates the scene at 1000 km (somewhat lower at sunspot minimum, higher at sunspot maximum) and shows a curious seasonal migration, piling up above the winter pole at the expense of helium above the summer pole. In the course of the year, the concentration of helium in the polar heterosphere varies by a factor of 4.5. The cause of this variation is largely unknown.

The uppermost part of the atmosphere consists of atomic hydrogen. While the production rate of hydrogen is fairly constant, its escape varies greatly with the exospheric temperature. The result is that there is much more hydrogen at sunspot minimum than at sunspot maximum; the difference can reach three orders of magnitude. Large seasonal-latitudinal asymmetries in the distribution of geocoronal hydrogen seem to be indicated by Lyman-alpha observations and may be expected also as a consequence of the so-called "polar wind," although a quantitative evaluation of these variations does not seem to be possible at this time.

CHAPTER IV. THE STABLE AURORAL RED ARC OF OCTOBER 31/ NOVEMBER 1, 1968 AND ITS INTERACTION WITH THE NEUTRAL ATMOSPHERE

By

R. G. Roble
National Center for Atmospheric Research,[1]
Boulder, Colorado

A. Abstract

The stable auroral red arc that occurred on October 31/November 1, 1968 is examined. Satellite observations of electron temperature and the topside electron density structure are used to calculate the red arc according to the thermal conduction model of the red arc. In this model, energy from the magnetosphere flows in the electron gas along geomagnetic field lines into the ionosphere. This energy heats the ambient F-region electrons sufficiently to excite the oxygen atoms to the OI(^{1}D) level by collisional impact giving rise to the λ6300 emission characteristic of the arc. For the October 31/November 1, 1968 red arc, the calculated λ6300 emission rate, geographical position, and horizontal extent of the red arc are in agreement with the photometric data obtained by airglow observatories.

Almost all of the energy conducted into the red arc is ultimately transferred to the neutral gas through elastic and inelastic collisions. This energy drives a large thermal cell and the circulation extends the influence of the arc thousands of kilometers beyond the region of direct heat input. The calculated neutral gas temperature response to electron heating within the arc is small, in agreement with the λ6300 doppler temperature measurements made with the Fabry-Perot interferometer at the University of Michigan Airglow Observatory.

B. Introduction

The stable auroral red arc (SAR-arc) that occurs in subauroral latitudes during geomagnetic storms was first discovered in the night sky by Barbier in 1957 [IV-1, IV-2]. Since then the arcs have been

1. The National Center for Atmospheric Research is sponsored by the National Science Foundation.

observed on numerous occasions from various ground-based airglow observatories, Roach and Roach [IV-3] summarized the properties of the red arcs that were determined during the last solar cycle maximum. More recently, satellite data have added considerably to our understanding of the red arcs, and Nagy et al. [IV-4] discussed these observations and reviewed the various hypotheses that have been proposed to explain the red arc.

The purpose of this paper is to examine the October 31/November 1, 1968 red arc in detail. Satellite data on the electron temperature and electron density structure within the red arc are used to calculate the arc according to the thermal conduction model of Cole [IV-5]; that is, energy conducted in electron gas from the magnetosphere along the geomagnetic field into the ionsphere heats the ambient F-region electrons. The heated electrons, in turn, excite the OI(^{1}D) term of atomic oxygen by collisional impact giving rise to the λ6300 emission characteristic of the arc. The calculated λ6300 emission features are compared with ground-based photometric observations made at the same time and in the vicinity of the satellite crossing of the red arc. In addition, the position of the red arc is shown to lie on the same L-shell as the plasma pause, indicating that the energy source for the red arc is in a narrow region just inside the plasma sphere.

The response of the neutral atmosphere to electron heating within the red arc is calculated and compared to the λ6300 doppler temperature and doppler wind measurements made within the October 31/November 1, 1968 red arc. These measurements were obtained with the 15.24-cm (6-inch) diameter, high resolution Fabry-Perot interferometer at the University of Michigan Airglow Observatory [IV-6 - IV-8]. Finally, the overall relationship of the red arc to the geomagnetic storm is discussed.

C. Summary of Observed Features

The observed features of red arcs in general have been discussed by many authors [IV-1 - IV-5, IV-8, IV-9 - IV-20]. These features are briefly summarized below.

1. The red arcs are observed only during geomagnetic storms and, once they form, they are persistent, homogeneous, stable, and have a lifetime of the order of 10 hours or longer.

2. The predominant emission within the red arc is the OI (^{1}D - ^{3}P) atomic oxygen transition at λ6300 with no large enhancement of the other emission lines normally present in the nightglow and aurora [IV-21]. In particular, the OI (^{1}S - ^{1}D) atomic oxygen transition at λ5577 is not greatly enhanced within the red arc region. The excitation energies of the OI(^{1}D) and OI(^{1}S) terms are 1.96 eV and 4.17 eV,

respectively; therefore, the lack of λ5577 emission implies that a low energy source must be acting within the red arc. The arcs are subvisual with intensities ranging from a few Rayleighs to tens of kilorayleighs [IV-3, IV-4].

3. The red arcs occur in the midlatitudes (Fig. IV-1) and are several hundred kilometers wide in the meridional direction. Current evidence indicates that the arcs extend in longitude at least around the nightside of the globe [IV-12, IV-14, IV-22]. In their longitudinal extent the arcs are at times slightly tilted with respect to constant L-shells, and the arcs occur simultaneously in the northern and southern hemispheres approximately along conjugate latitudes.

4. The mean height of the peak λ6300 intensity occurs near 400 km [IV-18, IV-23 - IV-25]. A meridional cross section of the normalized λ6300 intensity contours for a typical red arc is shown in Figure IV-2 [IV-13].

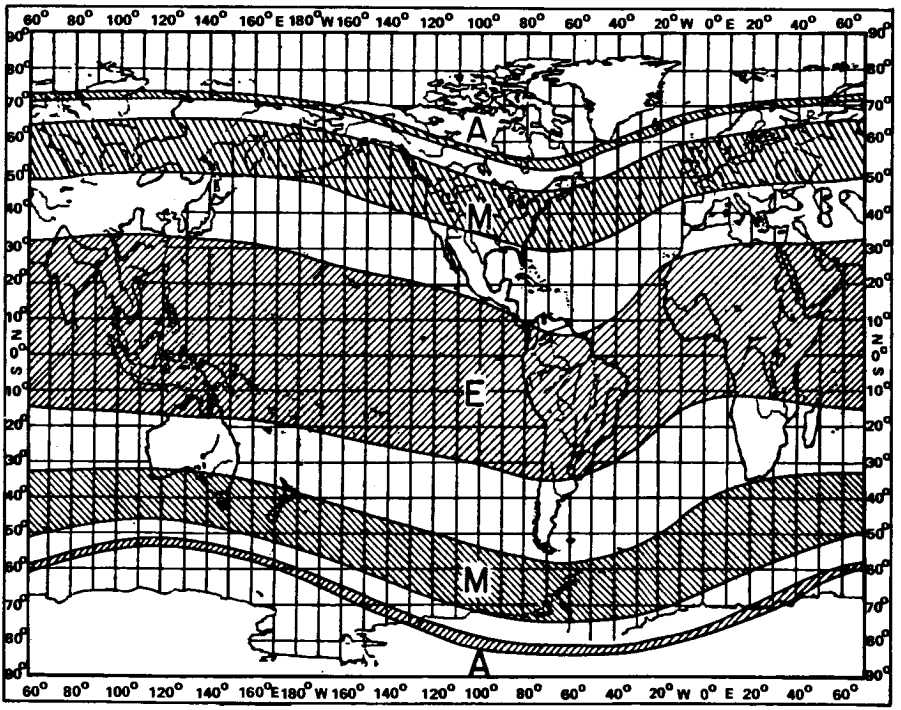


Figure IV-1. Zones of characteristic 6300 Å activity (the M region corresponds to a stable midlatitude red arc) [IV-3].

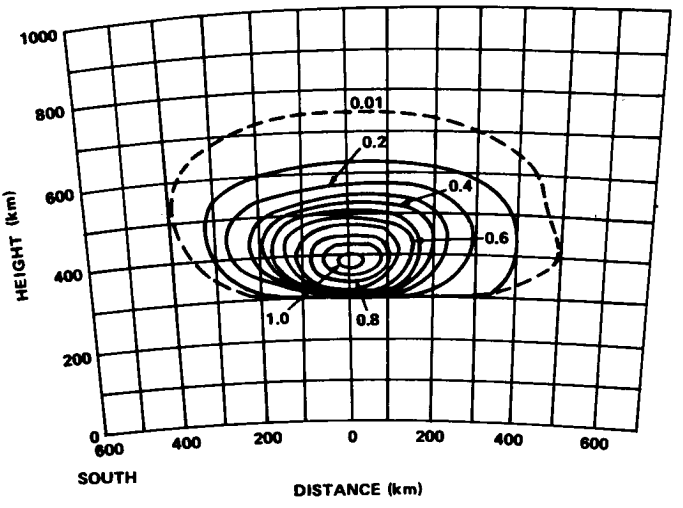


Figure IV-2. Normalized 6300 Å isophotal representation of an SAR-arc cross section [IV-13].

5. The arcs are observed only during periods of increased geomagnetic activity [IV-26], and they usually form after an aurora has occurred in the north [IV-10]. Recently, Hoch and Clark [IV-20] gave strong evidence that all red arcs are accompanied by poleward auroral displays and that they are spatially separate from poleward auroras. It is not clear, however, whether a red arc always accompanies a polar aurora.

6. The red arc appears to form on the same L-shell as the plasma pause [IV-27], and the red arc motion is apparently controlled by the movements of the plasma pause [IV-28].

7. Satellite observations show a significant increase in the electron temperature over the red arc region [IV-29, IV-30]. King and Roach [IV-11] first showed that the electron density decreased within the red arc region. More recent measurements by the Alouette 2 topside sounder show that a significant electron density depression exists within the red arc and that the depression is associated with the enhanced electron temperatures [IV-30 - IV-32].

D. Excitation Mechanisms

The four hypotheses that have been proposed as the excitation mechanism for the stable auroral red arc are (1) the dissociative recombination hypothesis [IV-11], (2) the electric field hypothesis,

[IV-15, IV-33], (3) the soft electron flux hypothesis [IV-34] and (4) the thermal conduction hypothesis [IV-5]. The various hypotheses have been discussed by Walker and Rees [IV-35] and Nagy et al. [IV-4], and it is generally agreed that thermal conduction from the magnetosphere is the excitation mechanism for the red arc. Roble et al. [IV-30, IV-36] used satellite data on the electron temperature and topside electron density measured over several red arcs and calculated each of the arcs according to the thermal conduction model. In each case the calculated λ6300 emission rate was in general agreement with the ground-based λ6300 photometric data. They also show that the calculated energy flux, determined from the satellite data, is sufficient to excite and maintain the red arc at its observed λ6300 intensity.

Recently, Okuda et al. [IV-37] suggested that an electric field was responsible for the minor λ6300 intensity variations that were observed in the March 8-9, 1970 red arc. The evidence for an electric field acting within the red arc was based on the correlation between the ground-based magnetometer variations and the changes of λ6300 intensity. Roble and Dickinson[2] show that an electric field cannot be the sole excitation source for the red arc. However, in a red arc primarily excited by thermal conduction from the magnetosphere, a superimposed electric field of the magnitude calculated by Okuda et al. [IV-37] is capable of exciting measurable λ6300 intensity variations.

E. Observations

A large geomagnetic storm, as revealed by the increased magnetic index K_p shown in Figure IV-3, began at about 0000 Universal Time (UT) on October 31, 1968 and lasted through November 3, 1968. The periods during which red arcs were observed from various ground-based airglow observatories are indicated in Figure IV-3, along with the University of Michigan Airglow Station observing period. In addition, the times of the Alouette 2 satellite crossings over the red arc region before and during the occurrence of red arcs are shown. The electron temperature was measured at the satellite and a topside sounder determined the electron density structure to the F-2 peak; these observations have been discussed in detail by Roble et al. [IV-30]. We are concerned here with the satellite crossing that occurred during the October 31/November 1, 1968 red arc at 0754 UT on November 1, 1968. The electron temperature measured along the satellite path during this pass is shown in Figure IV-4. The contours of electron density to the F-2 peak determined from the topside sounder are shown later in Figure IV-7a. The altitude

2. Roble, R. G.; and Dickinson, R. E.: Time-dependent behavior of a stable auroral red arc excited by an electric field. Submitted to J. Atmosph. Terr. Phys., 1971.

of the satellite was approximately 1200 km, and the geographic positions of the satellite pass are shown in Figure IV-5. The positions of the peak λ6300 emission within the red arc, as determined from the University of Michigan Airglow Observatory (42 deg 16 min North, 83 deg 44 min West) and labelled AA in the figure, during the two periods 0750 to

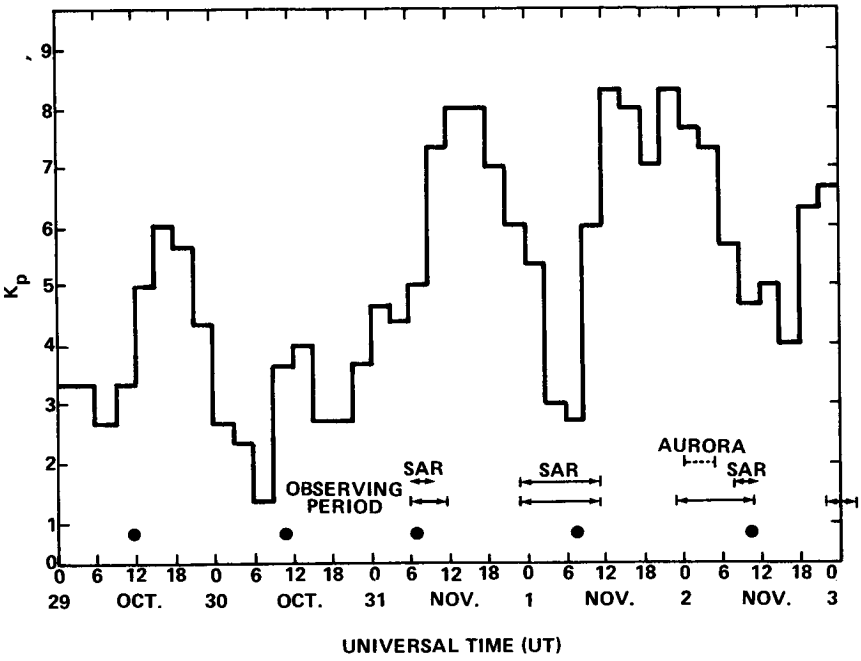


Figure IV-3. The magnetic index K_p as a function of time for the disturbed period October 30 through November 2, 1968 (the solid dots indicate the Alouette 2 crossing of the red arc region).

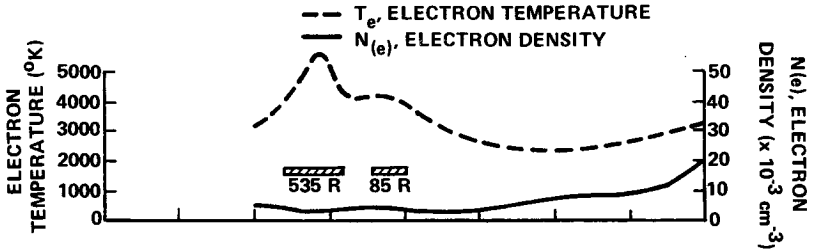


Figure IV-4. The electron temperature and density measured at the Alouette 2 satellite height as a function of magnetic invariant latitude.

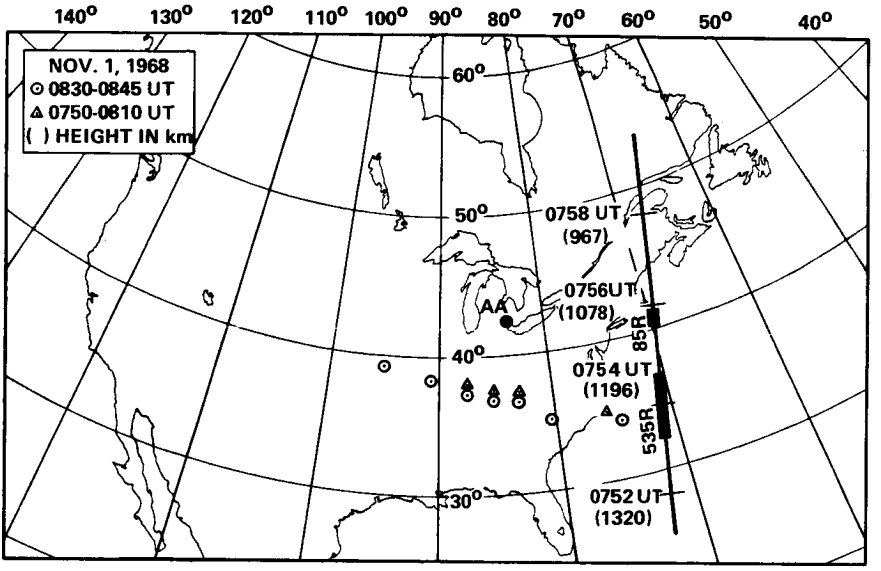


Figure IV-5. The geographic positions of the red arc of October 31/ November 1, 1968 determined from the University of Michigan Airglow Observatory at the time of the satellite crossing of the arc (the solid curve is the satellite path, and the universal time and height of the satellite are specified at various positions along the path; the heavy solid curve denotes the calculated position of the red arc and the corresponding numbers are the calculated λ6300 emission rates) [IV-30].

0810 UT and 0830 to 0845 UT are also shown. The position of the satellite crossing of the red arc is near the limit of the observing range of the airglow observatory. However, an extrapolation of the data points shows good agreement between the position of the red arc determined from the measured peak λ6300 emission and the peak electron temperature measured at the Alouette 2 satellite (Figure IV-4). The actual λ6300 photometric scans obtained from 0830 to 0845 UT are shown in Figure IV-6. The curves give the λ6300 emission rate as a function of zenith angle in the vertical circle surveys made by the λ6300 photometer and mirror scanning system [IV-6, IV-8]. The mirror scanning system was first positioned in azimuth and maintained at that position while a vertical circle scan was performed, going from +80 deg zenith angle in the north through the zenith to -80 deg zenith angle in the south. The mirror was then positioned to a new azimuth angle, usually 20 deg from the previous position, and the scan was repeated. The dashed curve, corresponding to an azimuth angle of 135 deg, is the scan that comes closest to the position of the satellite

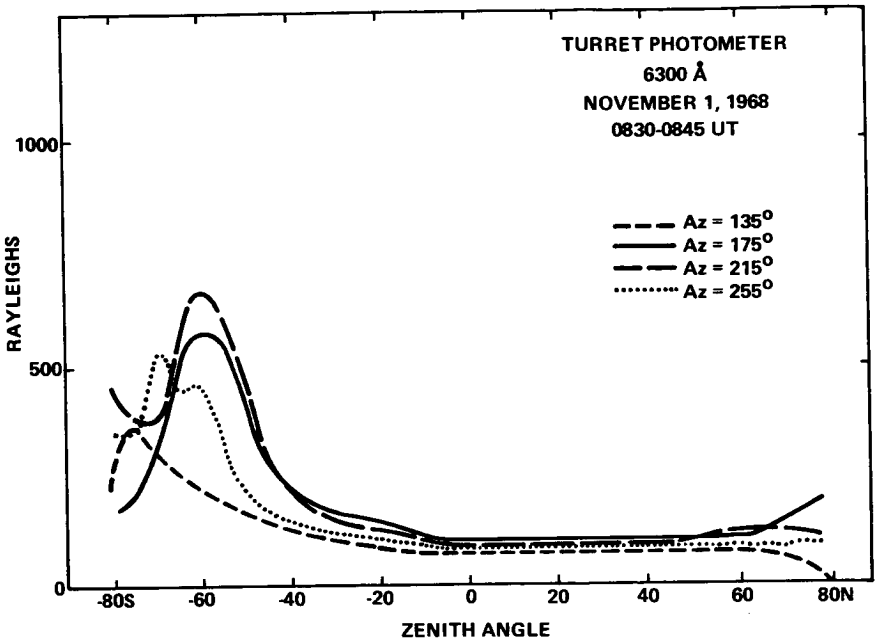


Figure IV-6. The λ6300 emission rate as a function of zenith angle for various azimuth angle positions (the vertical circle scan data were obtained by a turret photometer at the University of Michigan Airglow Observatory between 0830 and 0845 UT, November 1, 1968) [IV-6].

crossing of the red arc. The red arc is the λ6300 enhancement superimposed on the normal Van Rhijn intensity increase near the southern horizon. The sky conditions during the entire night of October 31/November 1, 1968 were generally clear, except for some autumn haze that persisted throughout the night. The moon set at 0730 UT; therefore, the scan shown in Figure IV-6 was not influenced by the moon's presence, which occurred in earlier scans [IV-8]. The measured peak λ6300 emission near the satellite crossing was approximately 400 Rayleighs. This value includes no correction for attenuation through the hazy sky. The measured λ6300 peak emission in the meridian plane was near 550 Rayleighs. Additional observations made on the October 31/November 1, 1968 red arc have been reported by Roble [IV-6], Hays et al. [IV-7], Roble et al. [IV-8], Roble et al. [IV-30], and Hoch and Clark [IV-20].

F. Analysis

Roble et al. [IV-36] used the electron temperature and topside electron density measurements of the Alouette 2 satellite as it passed over the September 28/29, 1967 red arc, and calculated the λ6300 emission in accordance with the thermal conduction model of Cole [IV-5]. The calculated λ6300 emission, photometric structure, height of the peak emission, and position of the calculated red arc were shown to be in general agreement with the observational data of the Fritz Peak Airglow Observatory. Roble et al. [IV-30], using the data obtained on several red arcs of varying intensities, found a consistent agreement between the calculated and observed λ6300 emission features. Their results indicate that thermal conduction of energy from the magnetosphere alone is sufficient to excite and maintain the red arc at its observed intensity.

Using the measured electron temperature at the satellite and the measured electron density to the F-2 peak, we solve the coupled electron-ion heat-conduction equations to obtain the electron temperature and λ6300 emission rate height profiles. For the nighttime red arc, the only energy source is heat, conducted in the electron gas from the magnetosphere into the ionosphere. The heating of the ambient F-region electrons is balanced by conduction and energy loss through elastic and inelastic collisions of the electrons with the ions and neutrals. The details of the interaction rates used in this analysis have been given by Roble [IV-6].

The one-dimensional, time-dependent, coupled, electron-ion heat-conduction equations are solved at various invariant latitudes along the satellite pass. The upper boundary condition on electron temperature is the measured electron temperature at the satellite height. In the ion gas, it is assumed that the energy flow at the upper boundary is zero. The lower boundary condition is that the electron, ion, and neutral gas temperatures are equal at 120 km. The initial electron and ion temperature height profiles are assumed, and the time-dependent coupled electron and ion heat-conduction equations are solved until a steady state is reached.

The calculated topside heat flow rate in the electron gas at the upper boundary is shown in Figure IV-7b. The region of the enhanced topside heat flow rate corresponds to the region of the enhanced electron temperature measured at the satellite. The positions of the enhancements also coincide with the position of the electron density depression shown in Figure IV-7a. The calculated latitudinal distribution of the λ6300 emission rate is shown in Figure IV-7b. The maximum λ6300 emission rate also occurs in the region of the enhanced topside heat flow rate. The calculated peak λ6300 emission rate is 522 Rayleighs, which is in agreement with the photometric observations. The contours of the calculated electron temperature within the red arc region are shown in

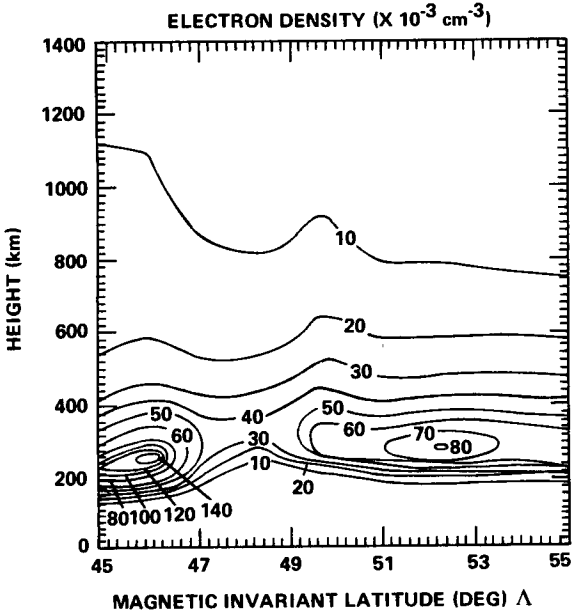


Figure IV-7a. Contours of electron density measured by the Alouette 2 topside sounder as it passed over the October 31/November 1, 1968 red arc [IV-30].

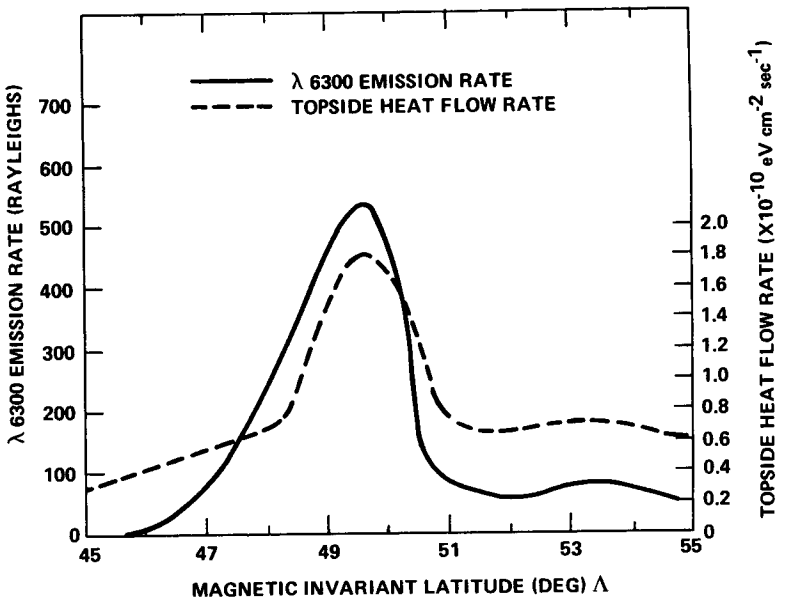


Figure IV-7b. The calculated topside heat flow rate and λ6300 emission rate as a function of the magnetic invariant latitude [IV-30].

Figure IV-8a. The enhanced electron temperature measured at the satellite penetrates deeply within the electron density depression. Because of the low plasma density, the hot electrons are not greatly cooled in elastic collision with the ions, and a high electron temperature extends down to the 400 km region where the λ6300 excitation is most effective. The calculated contours of the λ6300 volume emission rate for the red arc are shown in Figure IV-8b. The peak λ6300 volume emission rate occurs near 380 km, and the calculated contours are in reasonable agreement with the average contours of λ6300 emission determined by Tohmatsu and Roach [IV-13] from the data on several red arcs. The calculated position of the red arc is shown as the heavily shaded area along the satellite path in Figure IV-5. Although a direct measurement was not made at the satellite crossing of the red arc, an extension of the data points to the satellite crossing shows an agreement between the calculated and observed red arc positions. Thus, thermal conduction alone from the enhanced electron temperature measurements above the red arc to the ionosphere is sufficient to excite and maintain the red arc at its observed intensity.

G. Plasma Pause Position

It was shown in the previous section that an enhanced energy flow occurs in the region of the red arc as determined from the electron temperature measurements made by the Alouette 2 satellite. Nothing has been said about the origin of the energy flowing into the ionosphere. The narrow meridional extent of the red arcs indicates that the energy deposition must take place in a very narrow region just inside the plasma sphere. Chappel et al.[3] determined the position of the plasma pause during the October 31/November 1, 1968 red arc from the data obtained by the ion mass spectrometer aboard the OGO-V satellite. The observed position of the plasma pause and the position of the red arc are shown in Figure IV-9. In addition, the data they obtained on several red arcs showed the following common features: (1) red arcs are observed in the rapid-recovery phase of the geomagnetic storm in the 10 to 20 hours following the injection of the ring current particles, (2) the plasma sphere is sharply reduced in size during the geomagnetic storm to L-values characteristic of the red arc locations, and (3) the red arc is generally located in L-value near the position of the plasma pause.

Since Nagy et al. [IV-4] discuss the various proposed mechanisms for transferring the energy from the ring current particles to the ambient electrons in the vicinity of the plasma pause, the mechanisms will not be reviewed here. The close association between the red arc and plasma

3. Chappell, C. R.; Harris, K. K.; and Sharp, G. W.: OGO-V measurements of the plasma pause during observations of stable auroral red arcs. Submitted to J. Geophys. Res., 1971.

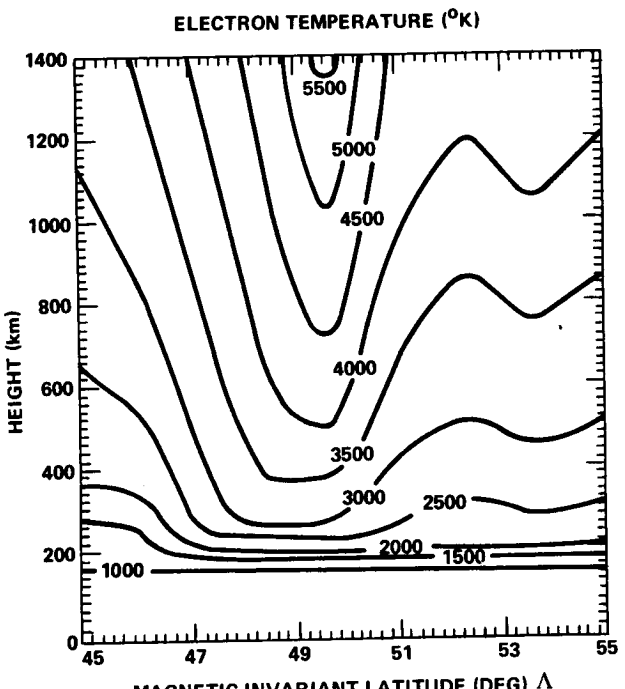


Figure IV-8a. The calculated contours of electron temperature as a function of magnetic invariant latitude [IV-30].

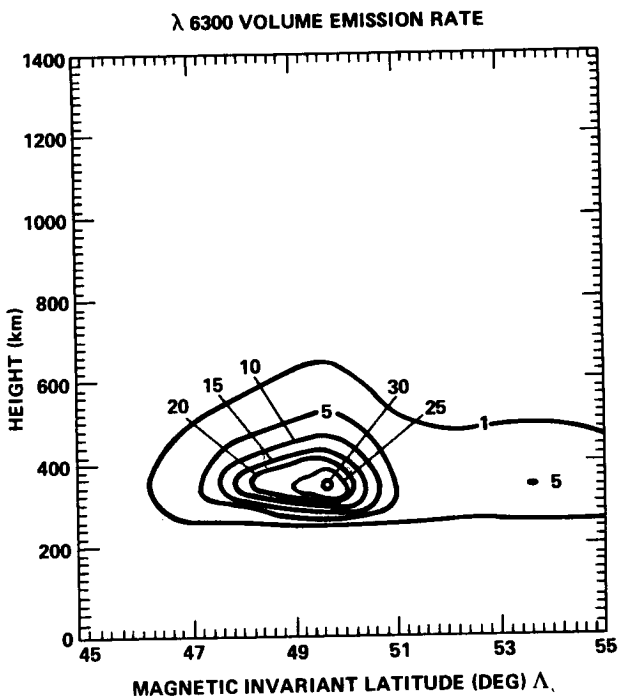


Figure IV-8b. The calculated contours of the λ6300 volume emission rate (photons cm^{-3} sec^{-1}) as a function of magnetic invariant latitude [IV-30].

pause positions, however, indicates that the motion and position of the red arc are probably controlled by the plasma pause movements. Also, the energy for the red arc appears to be derived from some form of interaction between the ring current particles and the ambient plasma in the vicinity of the plasma pause [IV-5, IV-16, IV-38 - IV-42].

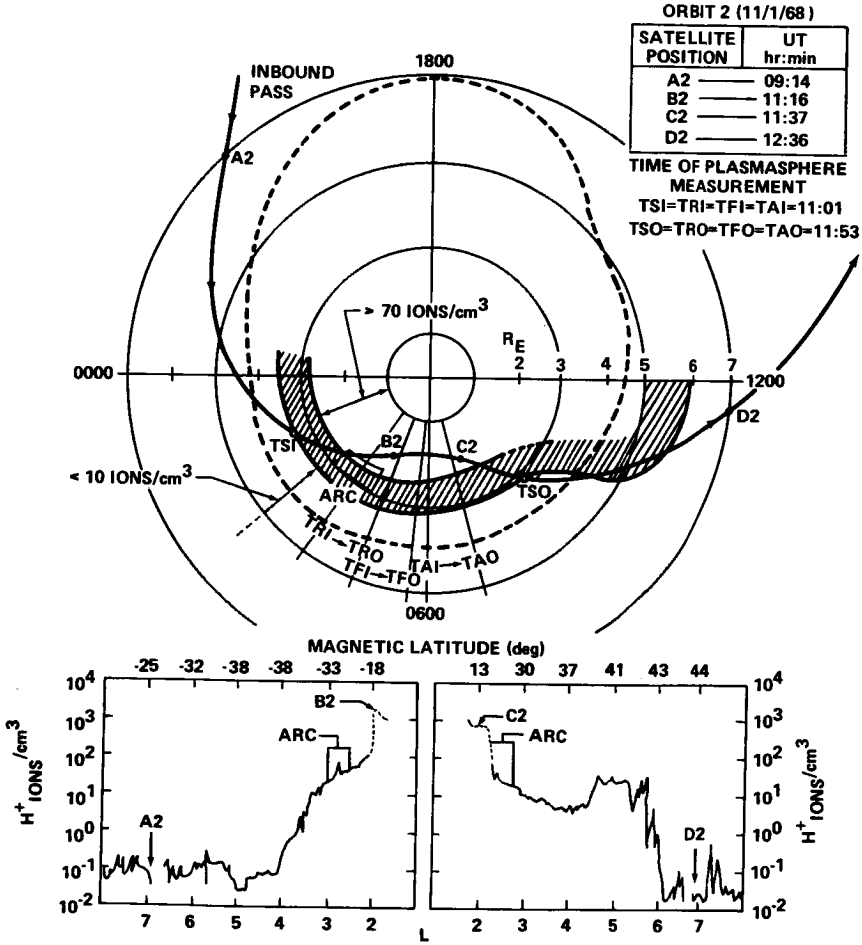


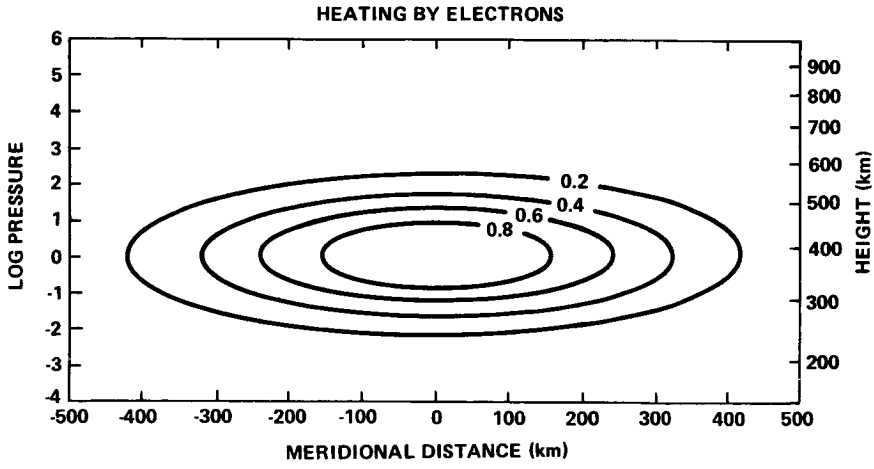
Figure IV-9. The OGO-V satellite pass of November 1, 1968 (the dashed line shows the average plasma pause location, and the cross hatched area approximates a portion of the plasma pause on this day; the rectangular-shaped area shows the position of a simultaneously observed red arc; and the insets at the bottom show the H^+ density profiles along the inbound and outbound orbits.

H. Atmospheric Response

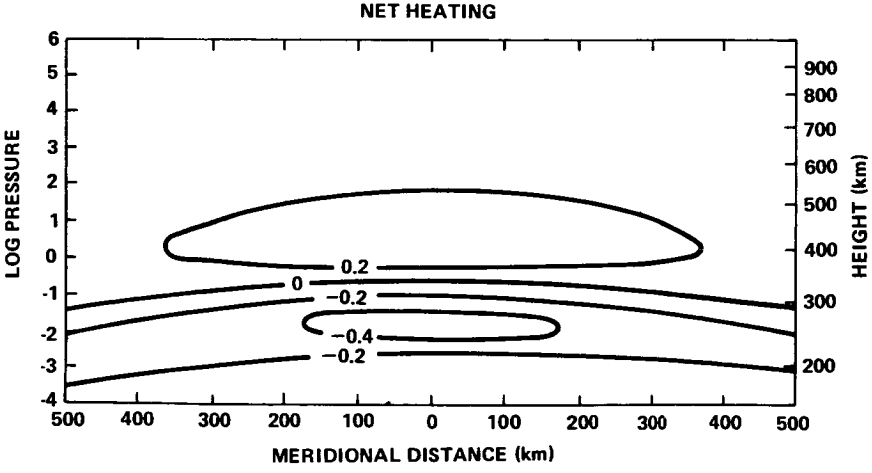
Of the energy conducted from the magnetosphere into the red arc, less than 0.5 percent is radiated as λ6300 emission, and very little energy is conducted in the electron gas out of the lower thermosphere. The bulk of the energy is ultimately transferred to the neutral gas through elastic and inelastic collisions of the electrons and ions with the neutrals. In this manner the stable auroral red arc acts as a heat source in the neutral thermosphere.

Roble and Dickinson [IV-43] examined the response of the neutral atmosphere to this heat source. They used a two-dimensional steady-state dynamic model of the neutral thermosphere to calculate the temperature perturbation and circulation pattern within the arc region. Figure IV-10a shows the normalized rate of heating per unit volume of the neutral gas by the electrons and ions within a red arc model. Figure IV-10b shows the sum of this heating rate and the rate of adiabatic (compressional) heating generated by the vertical motion field. The figures are scaled so that the maximum heating in Figure IV-10a is unity. Adiabatic expansion produces not only a reduction in the maximum heating but also gives rise to a significant region of cooling below the region of maximum heating.

Figures IV-11a and IV-11b show the temperature and horizontal wind perturbation in response to the net heating. The perturbation fields extend in a horizontal distance a factor of ten greater than the horizontal dimension of the actual source region. Within the red arc the lighter warm air produces a bulge in the pressure contours at a given height which pushes out the air horizontally until a balance is reached with viscosity and ion drag. By continuity, upward vertical velocities occur in regions of mass flux divergence and downward vertical velocities occur in regions of convergence. The net result is a large thermal cell with mass pulled in at low altitudes and raised in a relatively narrow "hot tower" where the heating is applied, with subsidence over a much more extended region. The maximum values of the neutral gas temperature increase and the maximum meridional velocity as a function of arc intensity are shown in Figure IV-12. The curve also relates the peak energy input for an integrated vertical column to the arc intensity. For the red arc intensities normally observed, the temperature and wind perturbation in the neutral atmosphere are small. Although these calculations apply to the red arc model described by Roble and Dickinson [IV-43], the results are general enough to estimate the maximum response for a specific red arc. For the October 31/November 1, 1968 red arc, the neutral heating from the 520-Rayleigh red arc causes a steady-state maximum temperature perturbation of approximately 20°K and a maximum meridional wind perturbation of 8 m sec^{-1}.

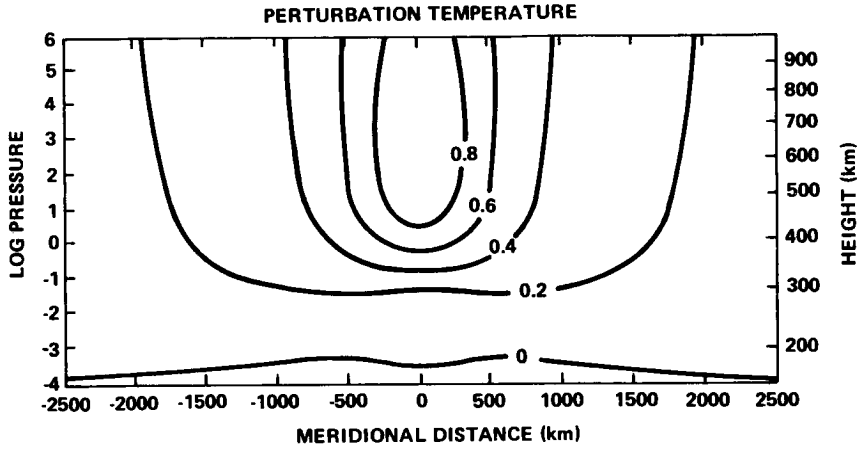


a. Heating by electrons and ions.

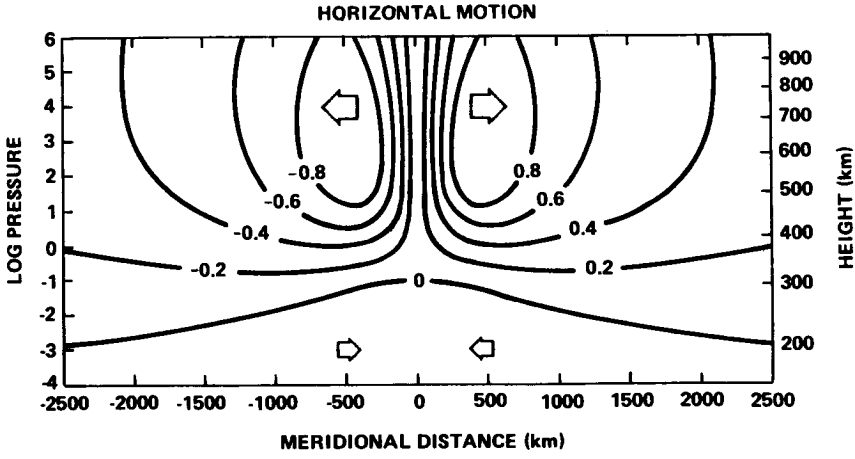


b. Sum of the nonadiabatic and the adiabatic heating.

Figure IV-10. The rate of heating of the neutral gas per unit volume within a red arc model normalized to a maximum value of one [IV-43].



a. Temperature perturbation.



b. Meridional wind perturbation.

Figure IV-11. Temperature perturbation and meridional wind perturbation for the red arc model (the maximum amplitudes are normalized to a maximum value of 1, and the maximum perturbation temperature and maximum amplitude of the meridional wind for a 500-Rayleigh red arc are 20°K and 8 m sec^{-1}, respectively [IV-43].

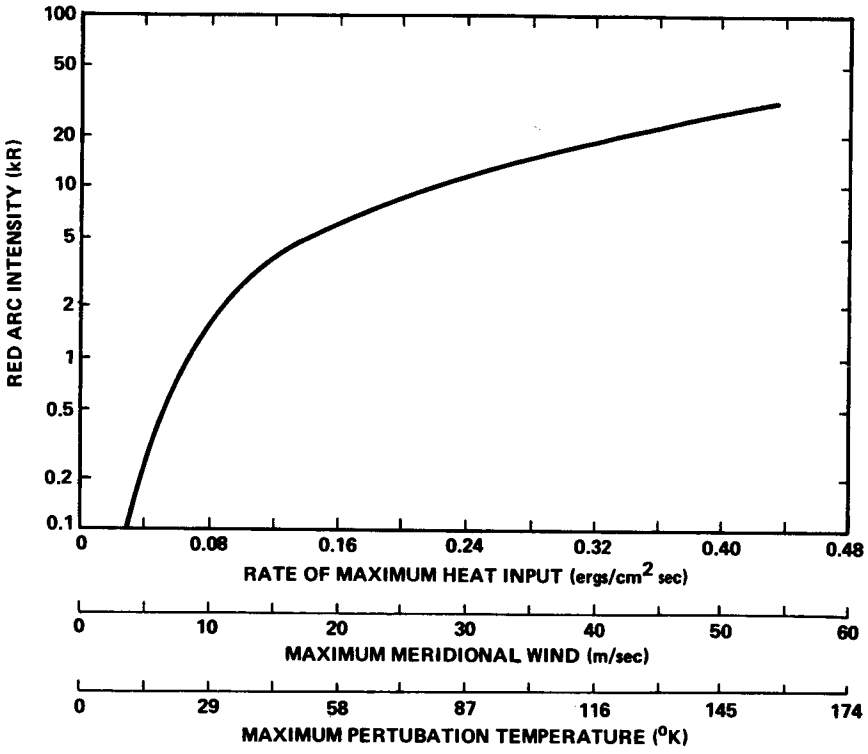


Figure IV-12. Conversion from maximum red arc intensity to maximum values of integrated energy input, temperature perturbation, and meridional velocity for the red arc model [IV-43].

I. λ6300 Doppler Temperature Measurements

The neutral temperature response within the October 31/November 1, 1968 red arc was measured by Roble et al. [IV-8, IV-36] with the 15.24-cm (6-inch) Fabry-Perot interferometer of the University of Michigan Airglow Observatory. The λ6300 doppler temperatures were made within and outside the red arc region (Fig. IV-13). The crosses in the figure are the doppler temperature measurements made on the night of October 30/31, 1968. The temperature remains nearly constant throughout the night at 1000°K. On the following night, the λ6300 doppler temperature measurements were considerably greater than the previous night and decreased during the night. The overall exospheric temperature variations during the geomagnetic storm period of October 30 through November 2, 1968 were reported by Hays et al. [IV-7]. The doppler temperature measurements inside and outside the red arc region showed no detectable temperature variations, indicating that the neutral

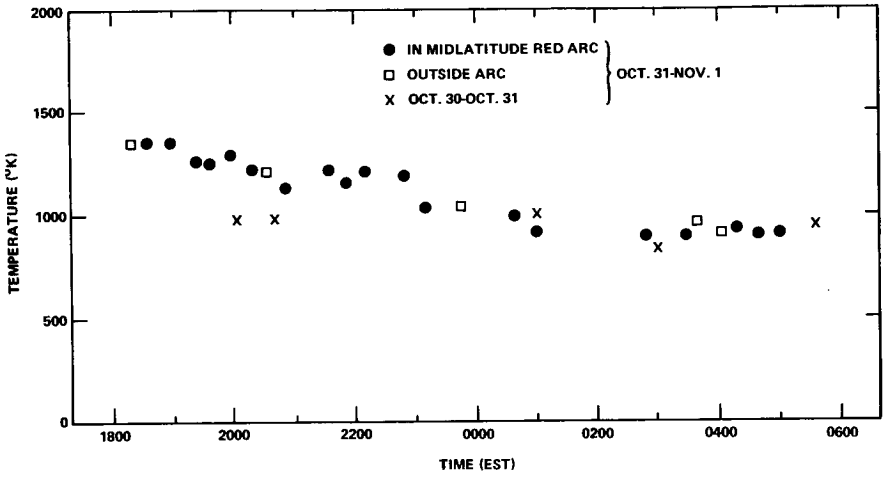


Figure IV-13. The λ6300 doppler temperature measurements during the magnetic storm of October 30 through November 1, 1968 [IV-8].

gas temperature increase within the arc is small. These results are consistent with the theoretical calculations of Roble and Dickinson [IV-43] indicating that the red arc has only a minor contribution to the overall thermospheric heating observed during the geomagnetic storm.

J. λ6300 Doppler Wind Measurements

In addition to the λ6300 doppler temperature measurements, the Fabry-Perot interferometer was used to measure the winds within the red arc region. The atmospheric winds in the vicinity of the red arc are determined from the measured doppler shift of two λ6300 fringe peaks: (1) one fringe profile measured in the zenith where the vertical component of the wind is known to be small [IV-43 — IV-46], and (2) one fringe profile measured at a high zenith angle where a component of the horizontal wind velocity lies along the line of sight of the instrument. The peak of the λ6300 fringe measured in the zenith is used to define the λ6300 line position in the absence of motion. Therefore, the shift of a λ6300 fringe peak at high zenith angles is a result of a doppler shift caused by the component of the horizontal wind along the line of sight of the instrument. The horizontal wind velocity is determined directly from the doppler relationship using the measured doppler shift in wavelength units and the zenith angle of the measurement [IV-47, IV-48]. The winds measured in the October 31/November 1, 1968 red arc are shown in Figure IV-14. Almost all the measurements were made in the southern sky in the region near the red arc, as observed from the

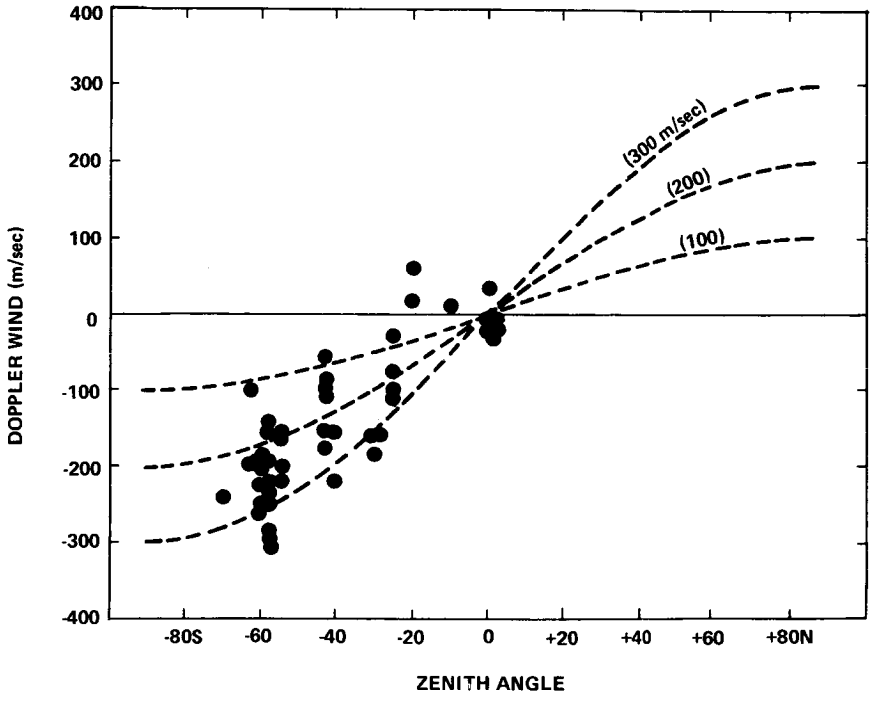


Figure IV-14. The λ6300 doppler wind measurements during the magnetic storm of October 30 through November 1, 1968.

University of Michigan Airglow Observatory [IV-8]. The measurements show a northerly wind component of the order of 250 to 300 m sec^{-1} in the altitude region of the maximum λ6300 emission near 400 km. These winds persisted throughout the night without any apparent changes in magnitude or direction.

The winds that were observed in the red arc region were not directly generated by the red arc. Roble and Dickinson [IV-43] show that almost all the energy deposited in the neutral atmosphere within the red arc is used to drive a large thermal cell; the calculated horizontal wind velocities within the thermal cell are small (approximately 8 m sec^{-1} for the 520-Rayleigh red arc). Therefore, the winds shown in Figure IV-14 are most likely a combination of the diurnal wind oscillation in the thermosphere and an enhanced wind blowing out of the auroral zone during a geomagnetic storm.

Horizontal winds of this magnitude flowing past the electron density depression, shown in Figure IV-7a, should experience a variation in the ion drag force. The ion drag on the neutral motions at F2-layer heights is

directly proportional to the ion and, hence, electron concentration. Large scale winds flowing past the electron density depression should accelerate in regions where the ion drag is less than the mean value and decelerate in regions where the ion drag is locally greater. Consequently, a perturbation of the horizontal flow on the scale of the anomaly occurs. By continuity, vertical velocities and the resulting distribution of adiabatic warming and cooling produce a temperature and pressure perturbation that balance in part the ion drag. These perturbations then add to those caused by direct heating within the red arc. Dickinson and Roble [IV-49] examined the atmospheric response resulting from the interaction of the mean winds and the ionization anomaly within the red arc. Their results show that the ion drag perturbation is mostly balanced by a back-pressure generated within the electron density depression. The resulting temperature perturbation caused by the adiabatic warming and cooling is small and of comparable magnitude to the temperature perturbation caused by direct heating.

K. Conclusions

During the geomagnetic storm of October 29 through November 2, 1968, significant energy was transferred to the magnetosphere from the solar wind. Part of this energy was immediately transferred to the lower atmosphere in the aurora. Another more gradual transfer of energy from the magnetosphere to the ionosphere occurred in the stable auroral red arc.

The red arc appears to be the sink for the energy causing the D_{st} geomagnetic variation during the storm [IV-4, IV-5, IV-16, IV-38 — IV-41]. Energy from the energetic ring current particles causing the geomagnetic perturbation is transferred to the ambient electrons in the region of the plasma pause. The exact mechanism by which this transfer takes place is not yet clear, but the relationship between the red arc and plasma pause positions is established [IV-27, IV-28, IV-41]. The energy that is transferred to the ambient electrons at the plasma pause is then conducted in the form of heat along the geomagnetic field lines into the ionsphere and heats the ambient F-region electrons. These heated electrons, in turn, excite the (^{1}D) term of atomic oxygen by collision, which gives rise to the λ6300 emission characteristic of the red arc. The lifetime of the red arc is of the order of 10 hours or longer, and it is persistent and stable during that time. Thus, the red arc is a manifestation of a slow release of energy that is stored in the magnetosphere during a geomagnetic storm. An energy balance calculation of a geomagnetic storm conducted by Cole [IV-5] showed that a slow release of the energy stored in the magnetosphere ring current is sufficient to sustain the red arc during its observed lifetime.

The bulk of the energy conducted into the ionosphere by the electrons and ions is ultimately transferred to the neutral gas through elastic and inelastic collisions by the electrons and ions with the neutrals. The red arc, therefore, acts as a heat source in the neutral thermosphere during geomagnetic storms. Roble and Dickinson [IV-43] have shown that the pressure forces originating from the red arc heating drive a large thermal cell with upward motion in the red arc region and a slow subsidence over a much greater area outside the arc. The adiabatic warming and cooling by these motions significantly reduce the neutral temperature within the red arc compared to the increase obtained in the absence of motion, but the warming and cooling enlarge the horizontal scale of the region of temperature increase. For the red arc intensities normally observed, the temperature and motion perturbations in the neutral atmosphere are small. Thus, the red arc gives only a minor contribution to the overall thermospheric heating that is observed during a geomagnetic storm.

L. References

IV-1. Barbier, D.: ' L' activité aurorale aux basses latitudes. Ann. Geophys., vol. 14, 1958, p. 334.

IV-2. Barbier, D.: ' L' arc auroral stable. Ann. Geophys., vol. 16, 1960, p. 544.

IV-3. Roach, F. E.; and Roach, J. R.: Stable 6300 Å auroral arcs in mid-latitudes. Planet. Space Sci., vol. 11, 1963, p. 532.

IV-4. Nagy, A. F.; Roble, R. G.; and Hays, P. B.: Stable mid-latitude red arcs: observations and theory. Space Sci. Rev., vol. 11, 1970, p. 38.

IV-5. Cole, K. D.: Stable auroral red arcs, sinks for energy of D_{st} main phase. J. Geophys. Res., vol. 70, 1965, p. 1689.

IV-6. Roble, R. G.: A Theoretical and Experimental Study of the Stable Mid-Latitude Red Arc (SAR-arc). Ph. D. Thesis, University of Michigan, 1969.

IV-7. Hays, P. B.; Nagy, A. F.; and Roble, R. G.: Interferometric measurements of the 6300 Å doppler temperature during a magnetic storm. J. Geophys. Res., vol. 75, 1969, p. 4162.

IV-8. Roble, R. G.; Hays, P. B.; and Nagy, A. F.: Photometric and interferometric observations of a mid-latitude stable auroral red arc. Planet. Space Sci., vol. 18, 1970, p. 431.

IV-9. Duncan, R. A.: Photometric observations of subvisual red auroral arcs at middle latitudes. Australian J. Phys., vol. 12, 1959, p. 197.

IV-10. Roach, F. E.; and Marovich, E.: A monochromatic low-latitude aurora. J. Res. Natl. Bur. Std., NBS 63D, 1959, p. 297.

IV-11. King, G. A. M.; and Roach, F. E.: Relationship between red auroral arcs and ionospheric recombination. J. Res. Natl. Bur. Std., NBS 65D, 1961, p. 129.

IV-12. Roach, F. E.; Barbier, D.; and Duncan, R. A.: Observation of a 6300 Å arc in France, the United States, and Australia. Ann. Geophys., vol 18, 1962, p. 390.

IV-13. Tohmatsu, T.; and Roach, F. E.: The morphology of mid-latitude 6300 Å arcs. J. Geophys. Res., vol. 67, 1962, p. 1817.

IV-14. Marovich, E.; and Roach, F. E.: Distribution of latitude of red arcs. J. Geophys. Res., vol. 68, 1963, p. 1885.

IV-15. Megill, L. R.; and Carleton, N. P.: Excitation by local electric fields in the aurora and airglow. J. Geophys. Res., vol. 69, 1964, p. 101.

IV-16. Cole, K. D.: The D_{st} Main Phase and Certain Associated Phenomena. Physics of Geomagnetic Phenomena, vol. II, Academic Press, New York, 1967.

IV-17. Marovich, E.: Fritz Peak Observations of Stable Auroral Red Arcs, Summary 1955-1965. ESSA Tech. Report, IER 16-ITSA, vol. 16, 1966.

IV-18. Hoch, R. J.; Marovich, E.; and Clark, K. C.: Reappearance of a stable auroral red arc at midlatitudes. J. Geophys. Res., vol. 73, 1968, p. 4213.

IV-19. Ichakawa, T.; and Kim, J. S.: Observations of M arc at Moscow, Idaho, U.S.A. J. Atmosph. Terr. Phys., vol. 31, 1969, p. 547.

IV-20. Hoch, R. J.; and Clark, K. C.: Recent occurrences of stable auroral red arcs. J. Geophys. Res., vol. 75, 1970, p. 2511.

IV-21. Carleton, N. P.; and Roach, J. R.: Spectroscopic observation of a mid-latitude red auroral arc. J. Geophys. Res., vol. 70, 1965, p. 1262.

IV-22. Reed, E. I.; and Blamont, J. E.: OGO-4 Observations of the September 1967 M-Arc. Trans. A.G.U., vol. 49, 1968, p. 641.

IV-23. Roach, F. E.; Moore, J. G.; Bruner, E. C.; Cronin, H.; and Silverman, S. M.: The height of maximum luminosity in an auroral arc. J. Geophys. Res., vol. 65, 1960, p. 3575.

IV-24. Moore, J. G.; and Odencrantz, F. K.: The height and geographical position of the red auroral arc of April 1-2, 1960. J. Geophys. Res., vol. 66, 1961, p. 2101.

IV-25. Rees, M. H.: A method for determining the height and geographical position of an auroral arc from one observing station. J. Geophys. Res., vol. 68, 1963, p. 175.

IV-26. Rees, M. H.; and Akasofu, S. J.: On the association between subvisual red arcs and the $D_{st}(H)$ decrease. Planet. Space Sci., vol 11, 1963, p. 105.

IV-27. Carpenter, D. L.: Plasmapause Observations from OGO VLF Data Near the Time of Red-Arc Events. Paper presented at the April 1970 National American Geophysical Union Meeting, Washington, D. C.

IV-28. Glass, N. W.; Wolcott, J. H.; Miller, L. W.; and Robertson, M. M.: Local time behavior of the alignment and position of a stable auroral red arc. J. Geophys. Res., vol. 75, 1970, p. 2579.

IV-29. Norton, R. B.; and Findlay, J. A.: Electron density and temperature in the vicinity of the 29 September 1967 middle latitude red arc. Planet. Space Sci., vol. 17, 1969, p. 1867.

IV-30. Roble, R. G.; Norton, R. B.; Findlay, J. A.; and Marovich, E.: The Calculated and Observed Features of Several Stable Auroral Red Arcs. Paper presented at 52nd Annual A.G.U. Meeting, April 12-16, 1971, Washington, D. C.

IV-31. Norton, R. B.; and Marovich, E.: Alouette Observations Taken During a Middle-Latitude Red Arc. Proc. IEEE, vol. 57, 1969, p. 1158.

IV-32. Clark, W. L.; McAfee, J. R.; Norton, R. B.; and Warnoch, J. M.: Radio Wave Reflections from Large Horizontal Gradients in the Topside Ionosphere. Proc. IEEE, vol. 57, 1969, p. 493.

IV-33. Megill, L. R.; Rees, M. H.; and Droppleman, L. K.: Electric fields in the Ionosphere and the excitation of the red lines of atomic oxygen. Planet. Space Sci., vol. 11, 1963, p. 45.

IV-34. Dalgarno, A.: Corpuscular radiation in the upper atmosphere. Ann. Geophys., vol. 20, 1964, p. 65.

IV-35. Walker, J. C. G.; and Rees, M. H.: Excitation of stable auroral red arcs at sub-auroral latitudes. Planet. Space Sci., vol. 16, 1968, p. 915.

IV-36. Roble, R. G.; Hays, P. B.; and Nagy, A. F.: Comparison of calculated and observed features of a stable midlatitude red arc. J. Geophys. Res., vol. 75, 1970, p. 4261.

IV-37. Okuda, M.; Old, T.; and Kim, J. S.: Mid-Latitude Auroral Arcs of 6300 (OI) and the Concurrent Ionospheric Current System. Proc. Upper Atmospheric Currents and Electric Fields Symposium, ESSA Tech. Mem. ERLTM-ESL 12, edited by W. H. Campbell and S. Matsushita, 1970.

IV-38. Cole, K. D.: Magnetic storms and associated phenomena. Space Sci. Rev., vol. 5, 1966, p. 699.

IV-39. Cole, K. D.: Magnetospheric processes leading to mid-latitude auroras. Ann. Geophys., vol. 26, 1969, p. 187.

IV-40. Cornwall, J. M.; Coroniti, F. V.; and Thorne, R. M.: Turbulent loss of ring current protons. J. Geophys. Res., vol. 75, 1970, p. 4699.

IV-41. Hoch, R. J.; and Smith, L. L.: Association of SAR-arcs and the Ring Current. Trans. A. G. U., vol. 11, 1970, p. 791.

IV-42. Coroniti, F. V.; Cornwall, J. M.; and Thorne, R. M.: A Theory for Mid-Latitude Red Arcs. Trans. A. G. U., vol. 11, 1970, p. 791.

IV-43. Roble, R. G.; and Dickinson, R. E.: Atmospheric response to heating within the stable auroral red arc. Planet. Space Sci., vol. 18, 1970, p. 1489.

IV-44. Dickinson, R. E.; and Geisler, J. E.: Vertical motion field in the middle thermosphere from satellite drag densities. Mon. Wea. Rev., vol. 96, 1968, p. 606.

IV-45. Dickinson, R. E.; Lagos, C. P.; and Newell, R. E.: Dynamics of the neutral gas in the thermosphere for small Rossby number motions. J. Geophys. Res., vol. 73, 1968, p. 4299.

IV-46. Rishbeth, H.; Moffett, R. J.; and Bailey, G. J.: Continuity of air motion in the mid-latitude thermosphere. J. Atmosph. Terr. Phys., vol. 31, 1969, p. 1035.

IV-47. Armstrong, E. B.: Doppler shifts in the wavelength of the OIλ6300 line in the night airglow. Planet. Space Sci., vol. 17, 1969, p. 957.

IV-48. Hays, P. B.; and Roble, R. G.: A technique for recovering doppler line profiles from Fabry-Perot interferometer fringes of very low intensity. App. Opt., vol. 10, 1971, p. 193.

IV-49. Dickinson, R. E.; and Roble, R. G.: Response of the Neutral Thermosphere to Large-Scale Anomalies of Ionization. Paper presented at the COSPAR Fourteenth Plenary Meeting, June 17-July 2, 1971, Seattle, Washington.

M. Bibliography

Carpenter, D. L.: OGO-2 and 4 VLF observations of the asymmetric plasmapause near the time of SAR-ARC events. Submitted to J. Geophys. Res., 1971.

Hays, P. B.: and Roble, R. G.: Direct observations of thermospheric winds during geomagnetic storms. Submitted to J. Geophys. Res., 1971.

CHAPTER V. SOME ASPECTS OF ATMOSPHERIC WAVES IN REALISTIC ATMOSPHERES[1]

By

Richard S. Lindzen[2]
University of Chicago, Chicago, Illinois

A. Abstract

Atmospheric tides travelling from the mesosphere into the upper thermosphere are discussed. As a preliminary, an elementary discussion of internal gravity waves is presented. Tides excited by the absorption of extreme ultraviolet radiation within the thermosphere are described also. Theory suggests a much larger semidiurnal oscillation in the thermosphere than has been observed.

B. Introduction

Under the broad title of this chapter we shall deal with internal gravity waves, and we will concentrate on a particular example; i.e., atmospheric tides. The following problem will be addressed.

Atmospheric tides below 90 to 100 km have been observed, and inviscid theory is quite successful in describing the observations. The tides are treated as rotationally modified internal gravity waves excited by the daily variations in sunlight as absorbed by ozone and water vapor. These tides are extremely strong in the 90 to 100 km region. Amplitudes may be as high as 50 °K in temperature and 100 m/sec in horizontal wind. There is a paucity of data in the region from 100 to 130 km, however, and the inviscid theories are inadequate in this region. Nevertheless, we know from satellite drag analyses that the atmosphere has large daily variations above 130 km. However, most attempts to explain the daily variations of the thermosphere have assumed that they are direct responses to the absorption of ultraviolet (UV) and extreme ultraviolet (EUV) radiation within the thermosphere itself. What, we may ask, has happened to the tides below 100 km in the region between 100 and

---

1. This work was supported in part by the National Science Foundation through grant GA25904.

2. Dr. Lindzen is an Alfred P. Sloan Research Fellow.

130 km? To answer this question I have studied the behavior of internal gravity waves in a viscous, thermally conducting atmosphere and have applied the results to tidal modes.

C. Internal Gravity Waves

Internal gravity waves have been discussed at great length in the literature [V-1 - V-3]. What I propose to do here is reproduce a simplified treatment that I developed earlier [V-4]. Internal gravity waves are waves whose restoring force is the buoyancy of displaced fluid elements in a stably stratified fluid.

Consider an element of fluid at some level z_o in a fluid where density ρ is decreasing with height at a rate $-\frac{d\rho}{de}$. The situation is depicted in Figure V-1a. The mass of the fluid element is

$$\delta m = \rho(z_o)\,\delta v \quad ,$$

where δv is the volume of fluid element. If we displace δm a small distance δs, it will be subject to a buoyancy force

$$-g\left\{\rho(z_o) - \left[\rho(z_o) + \frac{d\rho(z_o)}{dz}\,\delta s\right]\right\}\delta v$$

acting to return δm to z_o (g is the acceleration of gravity). Variations of δv because of compressibility have, for the moment, been neglected. The equation of motion for δm is

$$\rho(z_o)\,\delta v\,\frac{d^2\delta s}{dt^2} = g\,\frac{d\rho}{dz}\,\delta s\,\delta v$$

or

$$\frac{d^2\delta s}{dt^2} = \frac{g}{\rho(z_o)}\,\frac{d\rho}{dz}\,\delta s \quad . \qquad (1)$$

In a stably stratified, incompressible fluid, $\frac{d\rho}{dz} < 0$. Hence, equation (1) describes harmonic oscillation with a frequency ω_b given by

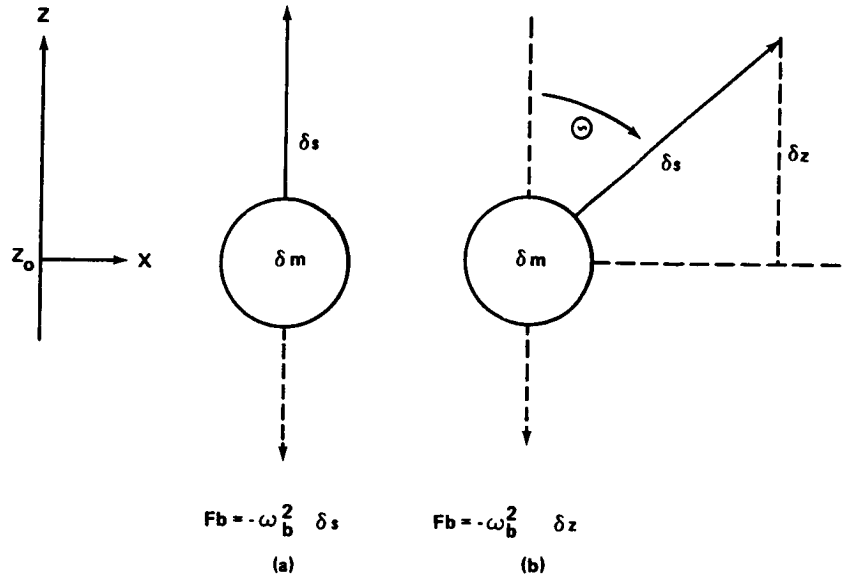


Figure V-1. Schematic description of fluid elements, their displacements, and buoyancy forces per unit mass [V-4].

$$\omega_b^2 = - \frac{g}{\rho(z_o)} \frac{d\rho}{dz} \quad , \tag{2}$$

where ω_b is known as the Brunt-Vaisala frequency. The effect of compressibility is to change our expression for ω_b to the following:

$$\omega_b^2 = \frac{g}{T} \frac{dT}{dz} + \frac{g}{c_p} \quad , \tag{3}$$

where T is the temperature of the ambient fluid.

Let us designate the buoyancy force per unit volume on a displaced fluid element at

$$F_b = - \omega_b^2 \ \delta s \tag{4a}$$

F_b is directed vertically. Now consider a fluid element that is somehow constrained to move at some angle θ with respect to the vertical

(Figure V-1b). The force exerted on this fluid element will be the projection of the buoyancy force,

$$F = -\omega_b^2 \, \delta z \cos\theta = -\omega_b^2 \cos^2\theta \, ds \quad , \tag{4b}$$

and the element will oscillate with a frequency ω, given by

$$\omega^2 = \omega_b^2 \cos^2\theta \quad . \tag{5}$$

A real fluid is, of course, continuous and cannot be thought of in terms of isolable elements. However, when we excite a real fluid at a frequency ω, where $\omega^2 < \omega_b^2$, the pressure forces do indeed constrain the fluid to move at an angle θ with respect to the vertical such that the projection of the buoyancy force corresponds to the restoring force appropriate to the frequency ω. The situation becomes more concrete when we consider waves excited by a corrugated bottom moving at a constant speed c (Figure V-2). The horizontal wavelength imposed by the corrugated bottom is L_H, which can also be expressed in terms of horizontal wave number $k = 2\pi/L_H$. Moving the bottom at a speed c gives rise to a local oscillation with frequency

$$\omega = kc \quad . \tag{6}$$

The angle θ is given by equation (5). If the fluid flow shown in Figure V-2 was plotted as a function of altitude at a fixed time and horizontal position, then we would see a sinusoidal structure with a vertical wavelength L_V (or vertical wave number $n = 2\pi/L_V$). Now,

$$\left(\frac{L_H}{L_V}\right)^2 = \left(\frac{n}{k}\right)^2 = \tan^2\theta \quad . \tag{7}$$

Combining equations (7) and (5) gives, approximately, the correct dispersion relation for internal gravity waves:

$$\left(\frac{n}{k}\right)^2 = \frac{\sin^2\theta}{\cos^2\theta} = \frac{1-\cos^2\theta}{\cos^2\theta} = \frac{1-\left(\frac{\omega}{\omega_b}\right)^2}{\left(\frac{\omega}{\omega_b}\right)^2} \tag{8a}$$

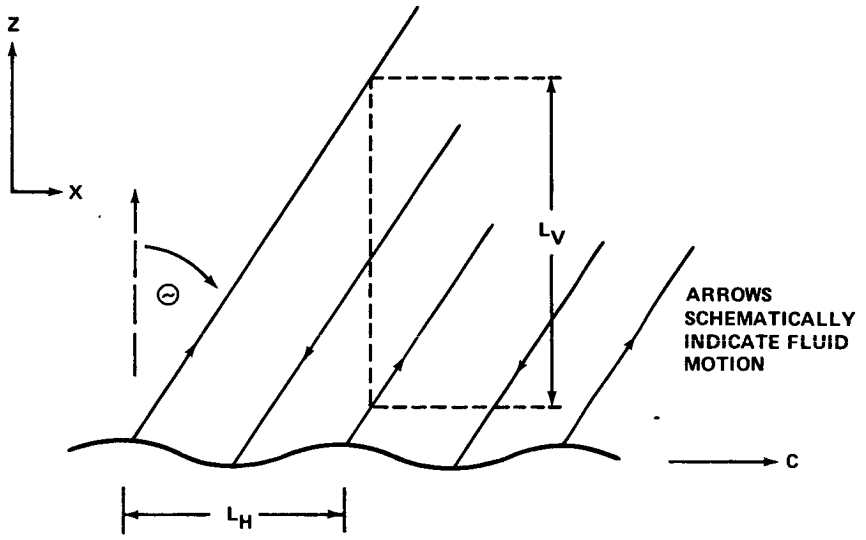


Figure V-2. Schematic description of motions excited in a stably stratified fluid by the motion (at speed c) of a corrugated bottom [V-4].

or

$$n^2 = \frac{1 - \left(\frac{\omega}{\omega_b}\right)^2}{\left(\frac{\omega}{\omega_b}\right)^2} k^2 \quad . \tag{8b}$$

Thus far, we are considering oscillations of the form

$$e^{i(\omega t + kx + nz)} \quad , \tag{9}$$

where

z = altitude,

x = horizontal distance,

and

t = time.

Such oscillations are waves that propagate upward with constant energy density in the absence of dissipation or mean flow. Since density is decreasing with height, the constancy of energy density consisting in terms like $\frac{1}{2}\rho u^2$, where u is the oscillatory horizontal velocity) requires that oscillatory fields increase in amplitude as $1/\rho^{1/2}$. The relations obtained thus far are appropriate only for isothermal atmospheres, where ω_b^2 is a constant. Because of the requirement of energy constancy, equation (9) must be replaced by

$$e^{z/2H}\ e^{i(\omega t + kx + nz)} \tag{10}$$

for an isothermal atmosphere, where

$$H = \frac{RT}{g} \quad , \tag{11}$$

R is the gas constant for air, and H is the atmospheric scale height. Because z dependence is no longer purely sinusoidal, equation (8b) must be slightly modified:

$$m^2 = \frac{1 - \left(\frac{\omega}{\omega_b}\right)^2}{\left(\frac{\omega}{\omega_b}\right)^2}\ k^2 - \frac{1}{4H^2} \tag{12}$$

Incidentally, we may see from Figure V-2 why upward energy flux for internal gravity waves implies downward phase progression. If the bottom is to do work on the fluid, it must push the fluid, in which case the fluid motions must tilt to the right of the vertical if c is positive. In this case a given phase appears, first aloft, and progresses downward.

Essential to the existence of internal gravity waves is the ability of fluid elements to retain their buoyancy (adiabaticity) and to move freely across isobars. The latter is inhibited by various frictional processes and also by rotation which tends to cause a fluid to move parallel to isobars. In a plane-rotating system, the dispersion relations for internal gravity waves become

$$n^2 = \frac{\omega_b^2 \left[1 - \left(\frac{\omega}{\omega_b}\right)^2\right]}{\omega^2 - f^2} k^2 - \frac{1}{4H^2} , \qquad (13)$$

where $f = 2\Omega$ and Ω = the rotation rate of the fluid. From equation (13) we see that the vertical propagation of an internal gravity wave requires

1. $f^2 < \omega^2 < \omega_b^2$

2. $k^2 > \frac{1}{4H^2} \frac{(\omega^2 - f^2)}{\omega_b^2 \left[1 - \left(\frac{\omega}{\omega_b}\right)^2\right]}$.

The requirement $\omega^2 > f^2$ is particularly important for atmospheric tides. On a rotating sphere we must consider $\Omega \sin \varphi$ (where φ is the latitude) to be the vertical component of the rotation vector, in place of Ω. Thus, f varies from zero at the equator to $2\pi/12$ hours at the poles. Hence, internal gravity waves with periods shorter than 12 hours can propagate vertically anywhere, but for longer periods internal gravity waves propagate vertically in regions increasingly confined to the vicinity of the equator. Thus, the diurnal tidal modes consist of vertically propagating modes confined primarily to the region between ± 30 deg latitude and exponentially trapped modes poleward of these latitudes.

Finally some comment is in order on the excitation of gravity waves (including tides). In our example, we considered excitation by a moving corrugated boundary. While pedagogically convenient, this was an obviously unrealistic example. Nevertheless, it has the element that is essential to any excitation; i.e., it causes time-varying vertical displacements of material surfaces. More practical ways of doing this are by fluid motions over stationary orography,[3] by explosions and volcanic eruptions, or by the daily variations in the absorption of sunlight by the atmosphere because of the rotation of the earth. The list is hardly exhaustive. In a stably stratified atmosphere, where dissipation is not dominant, any excitation mechanism with frequency components between ω_b and f will excite internal gravity waves.

What happens to internal gravity waves in atmospheres with viscosity and thermal conductivity increasing as $1/\rho$ is a complicated

3. By means of a shift of coordinates, this is equivalent to our idealized excitation; examples in nature are waves in the lee of wind flowing over mountains. Although mountains do not consist of infinite sinusoidal corrugations, their effects can be simulated by Fourier synthesis.

mathematical question. So too is the question of the effects of ion drag. Detailed treatments of these questions may be found in References V-5 through V-8. It will suffice here to sketch the results.

Almost any dissipative process whose effectiveness increases as $1/\rho$ tends to affect internal gravity waves in the same way. What happens to the wave depends on two parameters. First there is the ratio of the time scale for the wave $(1/\omega)$ to the time scale for dissipation (ρ/a),

$$X = \frac{a}{\omega \rho} \quad . \tag{14}$$

The dissipative process determines what a is. If we are considering viscosity,

$$a = \frac{4\pi^2 \mu}{L_V^2} \quad ,$$

where μ is the molecular viscosity. Since $X \propto 1/\rho$, it is an increasing function with altitude. The second important parameter is a ratio of atmospheric scale height to vertical wavelength,

$$\beta = \frac{2\pi H}{L_V} \quad . \tag{15}$$

If we are dealing with a nonisothermal atmosphere, then H is essentially the local scale height at the altitude where $X \sim 1$.

In terms of X and β the main effects of molecular viscosity and thermal conductivity are as follows:

1. The increasing dissipation with height serves as an inhomogeneity in the medium, which can cause downward reflection. The magnitude of the reflection is given by

$$|R| = e^{-\pi\beta} \quad . \tag{16}$$

L_V must be greater than approximately 120 km for this effect to be significant (assuming that H is approximately km).

2. For $\beta \lesssim 1$, wave amplitudes increase roughly as $e^{x/2}$ up to the vicinity of $X \sim 1$, asymptotically approaching a constant above this level, with little or no decrease in amplitude.

3. For $\beta > 1$, wave amplitudes increase roughly as $e^{X/2}$ up to the vicinity of $X \sim 1$, but then decrease considerably before asymptoting to a constant.

4. The effects of dissipation become important when $X \gtrsim 1$. Hence, for a given β, the greater ω is, the greater the height is at which dissipative effects set in.

5. The dominance of dissipation is associated with the constancy of both amplitude and phase with height.

6. Most of the above results apply to waves that propagate vertically in the absence of dissipation. For evanescent waves, increasing dissipation causes wave amplitudes to cease decaying with height and to approach a constant instead. This is accompanied by a change of phase.

Finally, a comment is in order on the effects of ion drag. Ion drag is centered in the F-region (about 250 km) where it has a time scale on the order of 1 hour. A priori one might expect ion drag to dominate the physics within the F-region of waves with periods in excess of 3 hours. However, the time scales for viscous diffusion are even shorter than those for ion drag, and the effect of the latter amounts to a reduction of no more than 40 percent in amplitudes.

D. Application to Atmospheric Tides

To apply the above to atmospheric tides, we must know something about tidal modes. These modes are three-dimensional internal gravity waves. Both their frequency and their zonal (east-west) wave number are imposed by excitation (rotation of an absorbing atmosphere in the sun's radiation). For each frequency and zonal wave number, there are an infinite number of tidal modes differing in their meridional (north-south) wave number. The structure of these modes is determined by the earth's geometry and rotation. The determination of these modes is a complicated matter described in detail in a monograph on atmospheric tides by Chapman and Lindzen [V-9]. As with the internal gravity waves discussed previously, the frequency and horizontal wave numbers of a mode determine its vertical structure, given the atmosphere's basic temperature distribution. Once we have the frequency and the vertical structure of a mode, we can evaluate X and β and hence the behavior of a mode in the viscous thermosphere.[4]

4. The application of results for internal gravity waves on a nonrotating plane to tidal modes involves certain unresolved problems that will not be discussed here. Discussions may be found in References V-5 through V-7.

In this discussion only the relevant results for the main migrating solar diurnal and semidiurnal modes will be cited. By migrating, we mean a mode that follows the sun. If f is a migrating tidal field, then

$$f = f(2\pi t + \varphi) \quad .$$

where

t = universal time in days

and

φ = longitude .

The diurnal part of f will behave as

$$e^{i(2\pi t + \varphi)} \quad ,$$

while the semidiurnal part will behave as

$$e^{2i(2\pi t + \varphi)} \quad .$$

Figure V-3 shows the symmetric semidiurnal modes. They are rather similar to Associated Legendre Polynomials. The symmetric diurnal modes are shown in Figure V-4. These fall into two categories; the first consists of modes concentrated within 30 deg of the equator which can propagate vertically, while the second consists of modes concentrated poleward of ±30 deg which are vertically trapped. The functions shown in Figures V-3 and V-4 are called Hough Functions. They are solutions of Laplace's Tidal Equation, and they are the natural expansion functions for the temperature, density, pressure, and vertical velocity fields associated with atmospheric tides. Somewhat different functions are used to expand the northerly and westerly velocity components. Table V-1 shows the properties of these modes relevant to their vertical structure in the presence of viscosity and thermal conductivity.

It is apparent from Table V-1 that none of the diurnal modes is effective in penetrating the thermosphere. In addition, it is shown in Reference V-7 that the main diurnal propagating mode suffers some measure of unstable breakdown near 90 km. Thus, there is no way in which diurnal oscillations excited by ozone and water vapor absorption

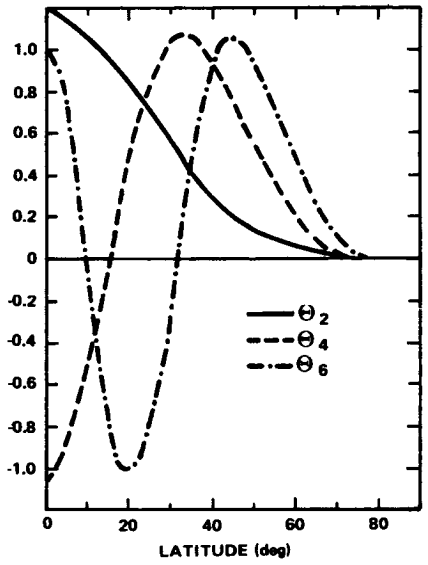


Figure V-3. Latitude distribution for the first three symmetric solar semidiurnal migrating Hough functions [V-9].

can significantly penetrate the thermosphere. The diurnal oscillations of the thermosphere must be excited in situ. The situation for semidiurnal tides is profoundly different. We see that the main semidiurnal mode continues to grow (approximately as $e^{x/2}$, where x is the height in scale heights) to very considerable altitudes and its amplitudes asymptote to constants with no decay. Thus, it appears that in addition to a diurnal oscillation, excited in situ, the thermosphere should also have a strong semidiurnal oscillation excited in the lower atmosphere.

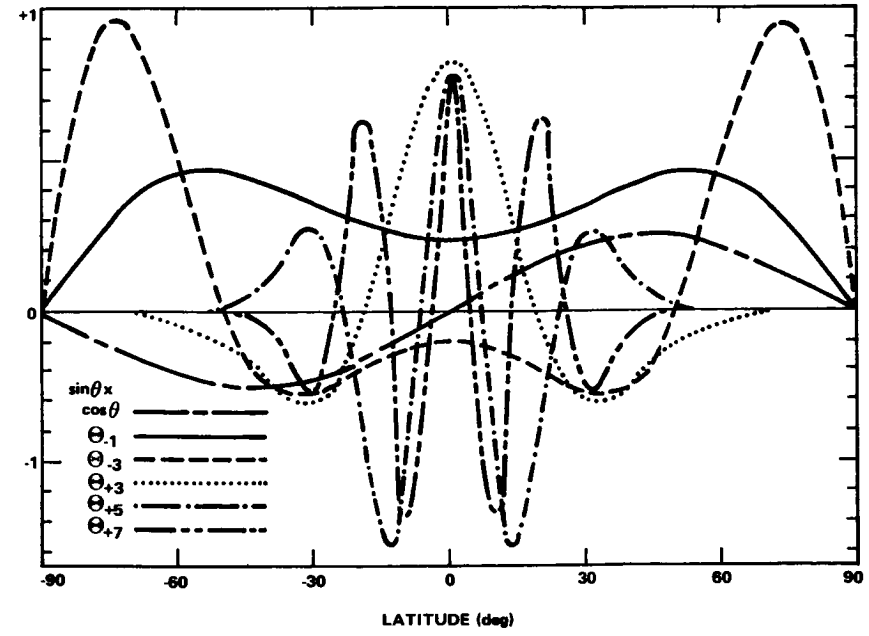


Figure V-4. Symmetric Hough functions for the migrating solar diurnal thermal tide; also shown is sin $\theta \times \cos \theta$, the most important odd mode.

TABLE V-1. TIDAL MODE PROPERTIES (all values apply where $X \sim 1$; $\overline{T}$ corresponds to the ARDC standard atmosphere).

	Mode	Altitude (km)	L_V (km)	H (km)	β
Semidiurnal	2, 2	145	120	16	0.84
	2, 4	125	45	11	1.6
Diurnal	1, 3	105	20	7.5	2.4
	1, -1	200	34a	24	-

a. Scale for vertical decay of amplitude.

E. Diurnal Oscillation in the Thermosphere

As indicated in the previous topic, the diurnal oscillation in the thermosphere must be excited in situ, primarily by EUV absorbed in the neighborhood of 150 km by N_2, O, and O_2. Therefore, the response of tidal-type gravity modes to such excitation has been calculated, including within the numerical calculations the effects of viscosity, thermal conductivity, and ion drag. The details are given in Reference V-6. Excitation due to EUV radiation was considered for which

$$\overline{\epsilon F_\infty} = 1.0 \text{ erg/cm}^2\text{/sec}$$

at noon. F_∞ is the incident solar flux, ϵ is the thermal efficiency of the absorbing processes, and the overbar indicates that the quantities have been integrated over the wavelength range of 220 to 900 Å. The diurnal component of the excitation was partitioned between the 1, 3 and the 1, -1 modes. Some of the results are shown in Figures V-5 and V-6. Figure V-5 shows the distributions with height of the amplitudes of the temperature oscillations associated with the 1, 3 and 1, -1 modes. Figures V-6 shows the phases associated with these modes. The amplitudes are appropriate to the equator. The total diurnal temperature oscillation results from the sum of the two modes. The oscillation at other latitudes may be calculated using Figure V-3. Roughly, we calculate a diurnal temperature oscillation with an amplitude of 200°K, with maximum temperature occurring at 1400 hours over the equator and between 1500 and 1600 hours at midlatitudes. While the phases are roughly in agreement with thermospheric observations (largely satellite drag data), the amplitude is substantially larger than that appropriate to recent Jacchia models. This is consistent with Hintereger's reduced estimate of solar

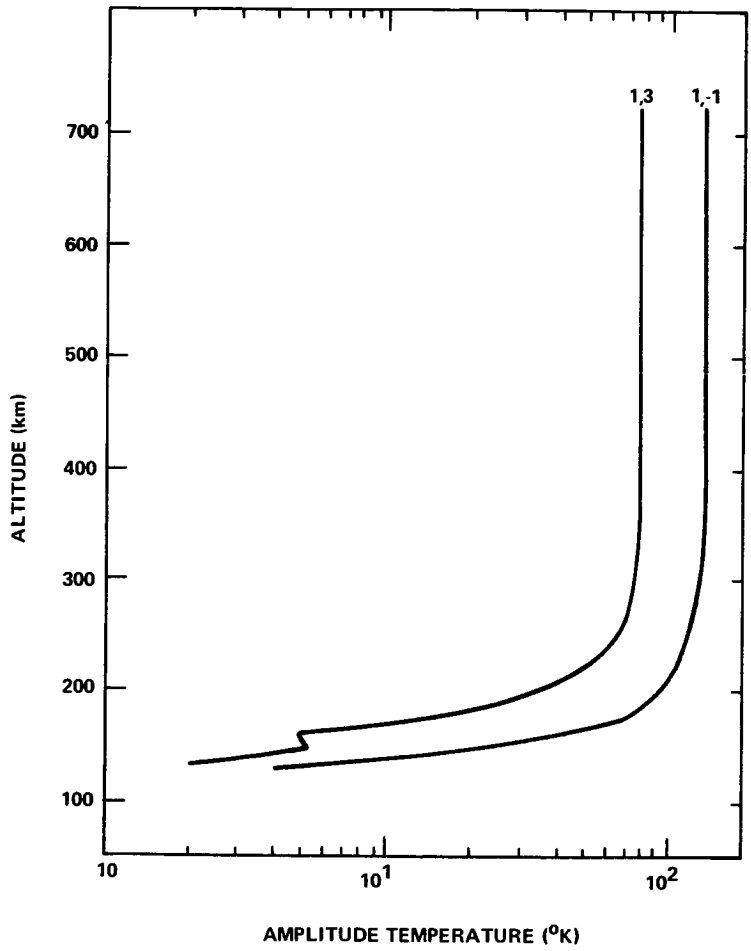


Figure V-5. Amplitudes of the diurnal temperature oscillations at the equator for the 1, 3 and 1, -1 modes as functions of height (excitation is a result of EUV absorption in the lower thermosphere) [V-6].

EUV flux and Hays' calculations of thermal efficiencies[5] which lead to an estimated noontime value of

$$\overline{\epsilon F}_{\infty} \approx 0.6 \text{ ergs/cm}^2\text{/sec}$$

5. Personal communications between the author and Mr. Hintereger and Mr. Hays established this result.

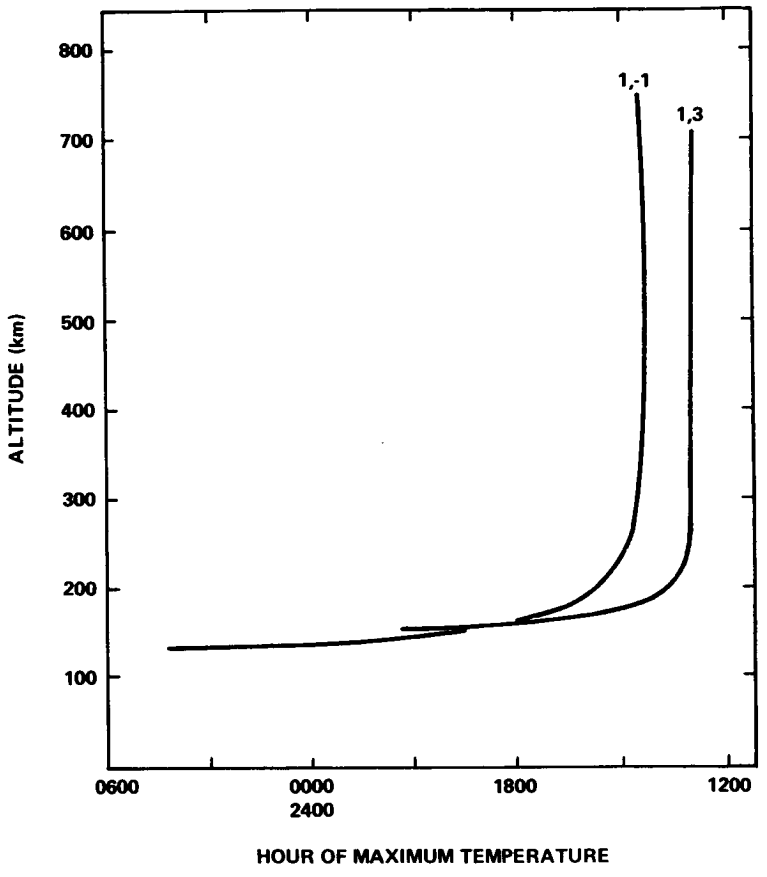


Figure V-6. Phases of the diurnal temperature oscillations at the equator for the 1, 3 and 1, -1 modes as functions of height (excitation is a result of EUV absorption in the lower thermosphere) [V-6].

for moderate solar activity. For such heating we would obtain an amplitude for the temperature oscillation of 120°K, which is substantially closer to Jacchia's models. Unfortunately, the lower value of $\overline{\epsilon F}_\infty$ is totally inadequate to maintain the observed mean temperature of the upper thermosphere. A consideration of the semidiurnal tide helps us out of the problem of maintaining mean thermospheric temperatures but leads to other difficulties.

F. Semidiurnal Oscillation in the Thermosphere

The results of section D suggest that semidiurnal tidal oscillations at approximately 90 km can effectively penetrate the upper thermosphere. Thus, it is of paramount importance that we know semidiurnal amplitudes

in the 90 to 100 km region. In Figure V-7 we show the height distributions of the semidiurnal zonal wind amplitude at various latitudes. Note the reduced growth of amplitudes in the upper mesosphere where, because of the low temperature and substantial lapse rate, the main semidiurnal mode becomes evanescent. Note also the increased importance of higher semidiurnal modes at higher latitudes. The predictions compare favorably with radio meteor data. Elford [V-10] found amplitudes of 10 to 15 m/sec at 90 km over Adelaide (35 deg south); Greenhow and Neufeld [V-11] found amplitudes of 15 to 20 m/sec at 90 km over Jodrell Bank (53 deg north); and Revah, et al. [V-12] found amplitudes of approximately 40 m/sec at 100 km over Garchy (47 deg north). Next we wish to calculate the upward extension of the semidiurnal tide into the thermosphere. The results are shown in Figures V-8 and V-9 where the height distributions of the amplitude and phase at the equator of the temperature oscillation associated with the main semidiurnal mode are shown. Results are shown for Schumann-Runge absorption, Schumann-Runge plus EUV absorption, and for all absorption including UV and near infrared absorption by O_3 and H_2O. Clearly, O_3 and H_2O absorption are the most important, and EUV is more important than Schumann-Runge absorption. Also shown are results for different thermospheric basic temperature distributions. The phase of the thermospheric semidiurnal temperature oscillation depends markedly on the basic temperature, though the amplitude does not. Our calculations suggest the presence of a semidiurnal temperature oscillation with an amplitude of 200°K. Since observations thus far do not suggest such a strong semidiurnal oscillation,

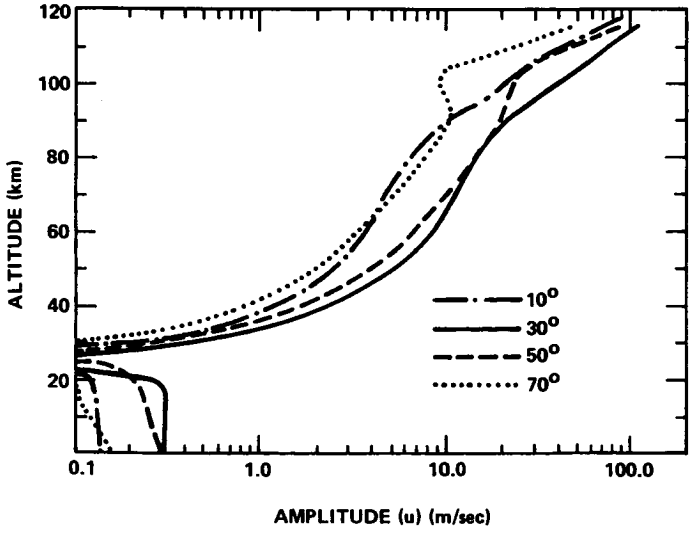


Figure V-7. Amplitude of the migrating semidiurnal oscillation in southerly wind as a function of height for different latitudes (excitation is a result of insolation absorption by O_3 and H_2O) [V-9].

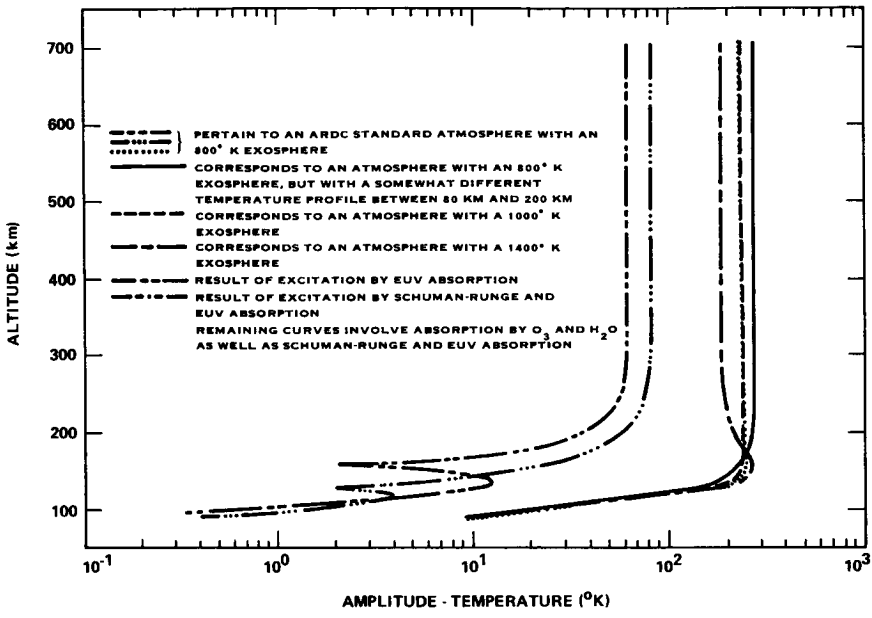


Figure V-8. Amplitude of the solar semidiurnal temperature oscillation over the equator as a function of height (the different curves correspond to different combinations of excitation and to different assumed basic temperature distributions) [V-6].

we are left in another quandary. However, the presence of the semidiurnal tide does provide us with a convenient solution to an earlier quandary. The upward propagation of the main semidiurnal mode is associated with a mean upward flux of energy amounting to approximately 0.3 to 0.4 ergs/cm^2/sec. If such a flux were a result of solar radiation, a noontime value of $\overline{\epsilon F_\infty}$ of approximately 1 erg/cm^2/sec would be needed. Such fluxes are small compared to UV fluxes. However, to heat the thermosphere, the flux must be deposited above approximately 120 km. As seen in Figure V-10 this is, in fact, the case for the mean energy flux carried by the semidiurnal tide. For comparison we also show the heating resulting from the deposition of a mean $\overline{\epsilon F_\infty}$ of 1 erg/cm^2/sec, which is currently believed to be inordinately high. In Figure V-11 the mean temperature that would result from the deposition of mean semidiurnal energy flux is shown. Clearly, the semidiurnal tide goes a long way toward relaxing the need for EUV heating. (Details of these calculations are given in Reference V-13.) However, the question remains as to why the semidiurnal oscillation has not been prominent in the data.

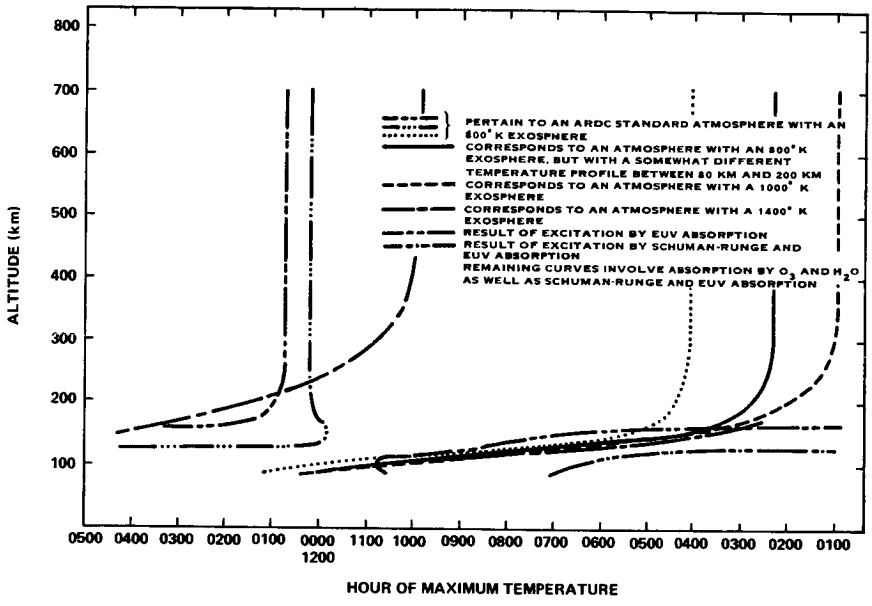


Figure V-9. Phase of the solar semidiurnal temperature oscillation over the equator as a function of height (the different curves correspond to different combinations of excitation and to different assumed basic temperature distributions) [V-6].

G. Remaining Problems and Possible Solutions

The above described calculations lead to what is, on the whole, a rather self-consistent picture of the upper thermosphere. The mean temperature is satisfactorily accounted for by the sum of EUV heating and heating resulting from the molecular dissipation of the mean energy carried by the semidiurnal tide. A relatively small EUV heating is sufficient to account for the observed diurnal oscillation in the thermosphere. However, a serious discrepancy remains between theory and observation for the semidiurnal oscillation. We have predicted a semidiurnal temperature oscillation of about 200 °K amplitude; Jaccia's models suggest no more than about 40°K. There are two obvious possibilities; the data are wrong or the theory is wrong. Both, in some measure, are true. Satellite drag data implicitly involve smoothing over 3 to 6 hours depending on orbit characteristics. Hence, such data do tend to suppress semidiurnal oscillations. In effect, satellite drag analyses attribute all drag to the perigee height; however, with a strong semidiurnal oscillation, much of the drag will come from regions away from perigee. Such variations are smoothed and perigee densities are overestimated. The exact extent of such errors has not yet been calculated. Similarly,

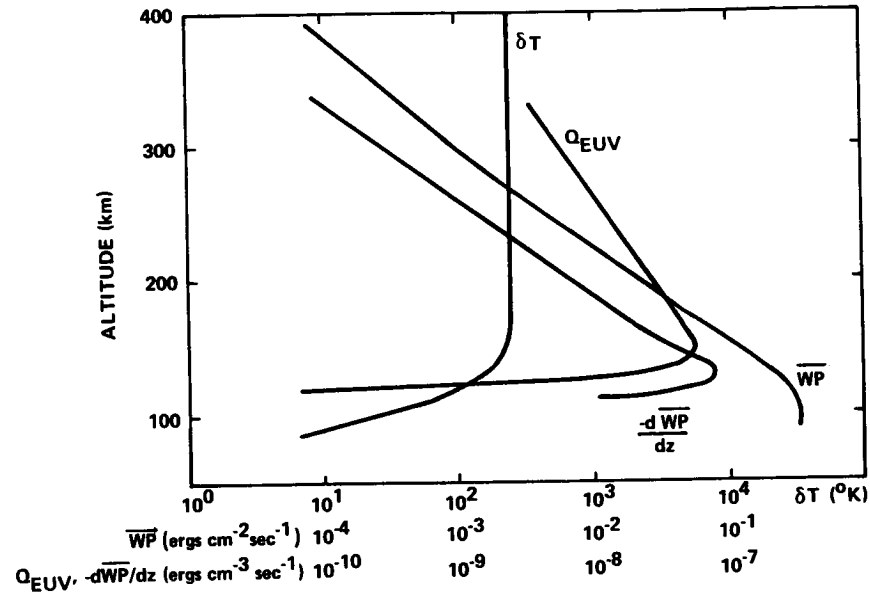


Figure V-10. Distributions with altitude of the amplitude of the solar semidiurnal temperature oscillation as excited by ozone and water-vapor insolation absorption (δT), the mean upward flux of energy resulting from the solar semidiurnal tide <WP>, the mean energy deposited by the solar semidiurnal ($-d\langle WP\rangle/dz$), and the mean energy deposited by absorption of EUV radiation – assuming $\overline{\epsilon F_\infty} = 1.4$ ergs/cm^2/sec at noon (Q_{EUV}) (values shown are appropriate to the equator) [V-13].

there are errors in the theory. An extensive discussion of such errors is given in Reference V-6. Such approximations as linearity and the use of "equivalent" gravity waves for the tidal modes can lead to errors of as much as 50 percent in the amplitude of the semidiurnal oscillation in the upper thermosphere.[6] (Observations support our calculations up to about 100 km.) Even so we would be left with an amplitude (i.e., half the range) of 100°K, which is larger than current data analyses suggest.

6. Surprisingly, ion drag does not seem capable of significantly reducing amplitudes.

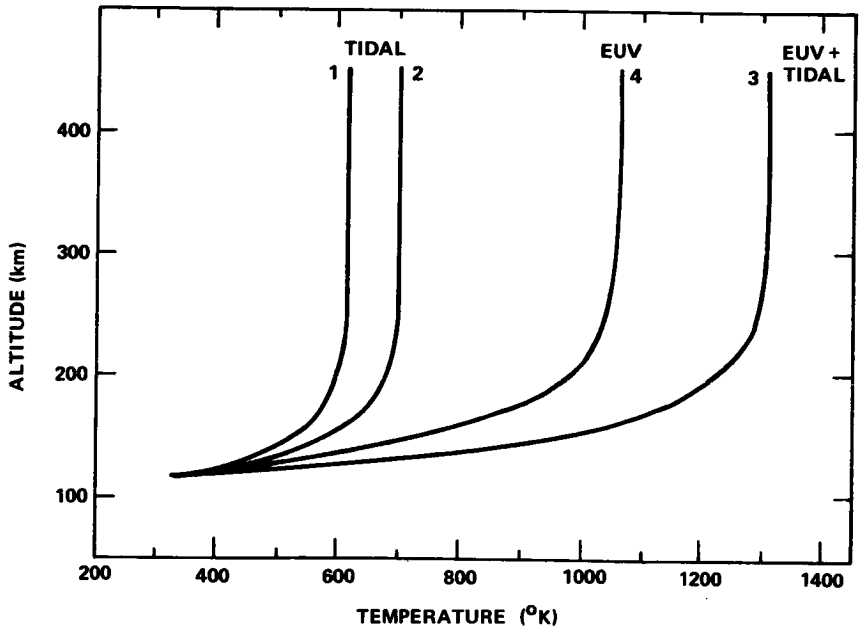


Figure V-11. Thermospheric temperature distributions resulting from the balance of thermal conduction with (1) the tidal heating shown in Figure 10, (2) one and one-third times the tidal heating shown in Figure 10, (3) the tidal heating used for curve 2 plus Q_{EUV} in Figure 10, and (4) Q_{EUV} alone (N.B. curves 1 and 2 are shown because the tidal energy flux is dependent on the details of the mean mesospheric temperature structure) [V-13].

H. References

V-1. Hines, C. O.: Internal gravity waves at ionospheric heights. Can. J. Phys., vol. 38, 1960, pp. 1441-1481.

V-2. Eckart, C.: Hydrodynamics of the Oceans and Atmospheres. Pergamon Press, New York, N. Y., 1960.

V-3. Jones, W. L.: Internal gravity waves. Handb. d. Physik., 1971.

V-4. Lindzen, R. S.: Tides and gravity waves in the upper atmosphere. Proceedings of the 4th ESRIN-ESROLAB Symposium, D. Reidel, Dordrecht, Holland, 1970.

V-5. Lindzen, R. S.: Internal gravity waves in atmospheres with realistic dissipation and temperature, 1, Mathematical development and propagation into the thermosphere. Geophys. Fluid Dyn., vol. 1, 1970.

V-6. Lindzen, R. S.: 3, Daily variations in the thermosphere. Geophys. Fluid Dyn., vol. 2, 1971.

V-7. Lindzen, R. S.; and Blake, Donna: Internal gravity waves in atmosphere with realistic dissipation and temperature, 2, Thermal tides excited below the mesopause. Geophys. Fluid Dyn., vol. 2, 1971, pp. 31-61.

V-8. Yanowitch, M.: Effect of viscosity on gravity waves and the upper boundary condition. J. Fluid Mech., vol. 29, 1967, pp. 209-231.

V-9. Chapman, S.; and Lindzen, R. S.: Atmospheric Tides. D. Reidel, Dordrecht, Holland, 1970.

V-10. Elford, W. G.: A study of winds between 80 and 100 km in medium latitudes. Planet. Space Sci., vol. 1, 1959, pp. 94-101.

V-11. Greenhow, J. S.; and Neufeld, E. L.: Winds in the upper atmosphere. Quart. J. Roy. Meteorol. Soc., vol. 87, 1961, pp. 472-489.

V-12. Revah, I.; Spizzichino, A.; and Massebeuf, Mme: Marée semidiurnelle et vents dominants zonaux measurés á Garchy (France) de Novembre 1965 á Avril 1966. Note Technique GRI/NTP/27, de Centre National d'Études des Telécommunications, Issy-les-Moulineaux, France, 1967.

V-13. Lindzen, R. S.; and Blake, Donna: Mean heating of the thermosphere by tides. J. Geophys. Res., vol. 33, 1970, pp. 6868-6871.

CHAPTER VI. UPPER ATMOSPHERIC SPECTROSCOPY[1]

By

A. E. S. Green, S. S. Prasad, and L. R. Peterson
Department of Physics and Astronomy
University of Florida, Gainesville, Florida

A. Abstract

Atomic physics is so intimately entwined with planetary aeronomy that progress in this geophysical-astronomical discipline is dependent in large measure upon progress in certain phases of atomic physics. In this article, we illustrate the interdependence by a discussion of photon, electron, and proton stimulation of the dayglow and the aurora. Calculations are given for the excitation of various states in N_2, O_2, and O as a result of electron energy degradation, assuming a gas mixture appropriate to 120 km. Emphasis is given to the $O_2(a^1\Delta_g)$ 1.27 μ emission and to the low energy region in areas where many competing mechanisms enhance or inhibit certain spectral emissions.

B. Introduction

The term "aeronomy" was introduced about 25 years ago to characterize the discipline concerned with the physical, chemical, and electromagnetic processes taking place in the upper atmosphere of the earth and planets. It is actually a very old discipline, and the study of auroras, an important facet of aeronomy, extends back to earliest historical times.

Rather than attempt to survey the full involvement of atomic physics with the atmospheric sciences [VI-1-VI-5], we will simply give several important illustrations. In particular, we will describe an approach to the calculations of natural upper atmospheric spectral emissions stimulated by photons, electrons, and protons originating from the sun and magnetosphere. These require as inputs a detailed knowledge of many atomic properties. Similar calculations are needed to explain atmospheric emissions induced by high altitude nuclear explosions. The results of such calculations might be compared with observations made with high resolution instruments from the ground or from rocket probes, satellites, and planetary fly-by vehicles. In the present work, we shall

1. This work was supported in part by AEC grant AT-40-13798 and NASA grant NGR-10-005-008.

illustrate our calculations by giving specific attention to the stimulation of low energy radiations.

C. Spectral Emissions Stimulated by Atomic Particles

Figure VI-1 is a block diagram giving a conceptual sequence of the steps in the calculation of the upper atmospheric spectral emissions resulting directly or indirectly from the solar input. Step 1 for photons requires precise spectroscopic measurement of the extreme ultraviolet and soft X-ray flux. Such data, which are now available from satellites, display large temporal variations associated with solar activity. The geometry problem for photons impacting upon the earth (step 2) is relatively simple except for grazing incidence at sunrise or sunset or in the polar regions. In such cases, instead of a simple secant law for the path length, one has a complicated absorption problem in which the spherical geometry and the altitude distribution of the atmospheric species interplay in a complicated way.

Steps 1 and 2 for electrons are more difficult to execute, and despite the expenditure of tremendous effort, electron flux measurements

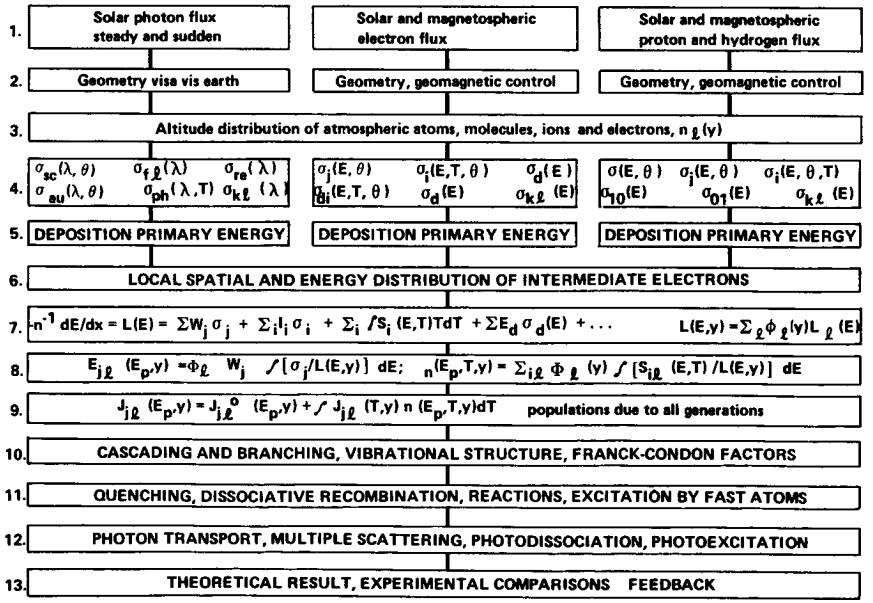


Figure VI-1. Block diagram of steps leading to a determination of spectral emissions stimulated by photons, electrons, and protons.

from rockets and satellites constantly bring surprises. Very intense fluxes, increasing as one goes below 1 keV, have been observed among trapped electrons and also among the electrons that are dumped into the auroral zone. The geomagnetic influence upon the passage of the electrons is exceedingly complicated, especially since the magnetic field changes with a geomagnetic disturbance. The same remarks apply to protons and hydrogen atoms associated with solar wind, solar flares, and the magnetosphere. Here it should be noted, however, that, unlike electrons, protons can undergo charge exchange and thus become decoupled from the magnetic field. Upon stripping at some later point, these fast moving particles again couple to the field. These charge changes cause a complicated diffusion of protons with respect to the field.

Step 3 in all channels involves the specification of the altitude number distribution $n_\ell(y)$ of the atomic, molecular, and ionic species in the atmosphere that are encountered by the primary particles. Step 4, for the case of photons, involves the totality of inelastic and scattering cross sections that control the fate of the primary photon of wavelength λ. Here, $\sigma_{sc}(\lambda, \theta)$ refers to the differential scattering cross section and determines the angle θ of the outgoing photon; $\sigma_{f_\ell}(\lambda)$ and $\sigma_{re}(\lambda)$ describe the various fluorescent and resonant processes; $S_{au}(\lambda, T)$ and $S_{ph}(\lambda, T)$ (differential with respect to secondary electron energy) provide for the production of Auger and photoelectrons with energy T; and $\sigma_{i,k}(\lambda)$ describes multiple ionization. Many higher order cross sections could be added to the list if required for special applications.

Step 4 for electrons of energy E depends on the cross sections $\sigma(E, \theta)$ for differential elastic scattering, $\sigma_j(E)$ for inelastic processes associated with the production of an excited state j, $S_i(E, T)$ describing ionizations leading to additional outgoing electrons with energy T, $S_{di}(E)$ and $\sigma_d(E)$ for dissociative ionization and excitation processes, and $\sigma_{i,k}(E)$ for multiple ionization. Again, several other processes could be included if needed. Step 4 for protons again involves the complete array of elastic, excitation, ionization, charge exchange, and other cross sections that characterize the interactions of incident protons and hydrogen atoms with the atmospheric species.

The arrays of cross sections illustrated in step 4 control the time, spatial, and energy history of primary particles. However, how to use such cross sections, even if they were known, is still a matter of uncertainty. Eventually, a Monte Carlo type of calculation might be most practical to carry out step 5, but we are currently leaning toward primarily simple techniques for scaling laboratory data.

The energy deposition of primary particles leaves a trail of excited atoms, molecules, and ions, as well as a host of electrons. These secondary electrons range in energy up to the kilo-electron-volt region, but they are predominantly below 100 eV and have a complex spatial distribution. To discuss the photon, electron, and proton problems in parallel, we will refer to electrons generated by the primary particles in each of the three channels as intermediate electrons. Therefore (step 6), we must again concern ourselves with details of all excitation and other inelastic electron induced processes coming from the intermediate electrons. Here, the problem is quite unconventional with respect to typical problems in pure atomic physics. Because of the failure of the Bethe-Born approximation and certain sum rules applicable at high energies, an analysis requires detailed characterizations for all inelastic cross sections, not just the specific ones related to spectral lines of interest.

One approximate method [VI-6-VI-10] for dealing with the energy deposition problem is to construct the loss function L(E) or stopping power, whose definition is indicated in step 7. The successive terms represent the loss contributions because of excitation of atomic or molecular states with excitation energies W_j, ionizations with the threshold energies I_i, production of intermediate electrons having a continuum of energy, dissociations with thresholds E_d, and any other significant loss processes consistent with the previously defined cross sections. For the lower energy electrons, vibrational excitation losses and, if the ambient electron density is high enough, electron-electron elastic losses become important. We must also compound the loss function for individual species to find the net loss function at various altitudes.

For most purposes we may assume that the intermediate electron loses its energy locally. For example, at 120 km, the number density is about 10^{11} cm^{-3} and L(E) varies from 10^{-4} eV cm^2 down to about 10^{-15} eV cm^2 for electrons between 10 and 100 eV. Thus, a typical 100-eV electron loses energy at a rate of about 1 eV per 10^4 cm and moves about 10 km before stopping. This contrasts with the sea level value, $n \approx 10^{19}$ with movement of about 10^{-2} cm, or a biological system, where $n \approx 10^{23}$ with only 10^{-6} cm movement.

The next step (step 8) is to compute $E_j(E,y)$, the primary electron energy transferred to the j-state, and to calculate the secondary electron spectral distribution $n(T,E,y)$ generated as a result of the primary degradation. Equivalently, the population $J_{j\ell}{}^0(E) = E_{j\ell}(E)/W_j$ may be obtained. This process must be carried out for each component of the intermediate electron spectrum. Once the complete secondary electron spectrum is known, the calculation of the energy transferred to the j-state can be iterated. Fortunately, this approach converges rapidly, and the contributions of the third and certainly the fourth generation are usually relatively insignificant. Alternatively, an integral equation (step 9) may

be used to give the total population $J_{j\ell}(E)$ of each ion or excited state from all the generations of degradation once the primary populations $J_{j\ell}{}^{0}(E)$ are known. The fast (10^{-8} sec) radiative decay of excited states is the concern of step 10, and this requires a knowledge of the entire array of transition probabilities. The slower time scale (seconds to hours) of the metastable states and the fate of the active atoms, ions, and free electrons are the concern of step 11. Finally, step 12 deals with the radiative transport of atmospheric emissions and includes the photochemical processes induced by the harder ultraviolet emissions.

The consequences of steps 1 through 12 are theoretical results that are to be compared with field observations (step 13). Where preliminary calculations of this type have been carried out [VI-6-VI-8, VI-10, VI-11], a factor-of-two consistency is usually realized between theory and experiment. However, when disagreements occur (often orders of magnitude) the feedback experience to date indicates that the culprit is about as likely to be an erroneous atomic property as an erroneous aeronomical property.

D. Energy Levels and Cross Sections

With the recent rapid advances in knowledge of atomic and molecular energy levels and electron impact cross sections, there is a continuous need for updating spectral emission calculations. We have attempted to do this, in connection with this survey, by proceeding upward from the lowest lying states. This affords us the opportunity to give particular attention to the 1.27-μ emission of O_2 ($^1\Delta_g$), which is of considerable current interest because of the unexpectedly large intensities that have been reported [VI-12-VI-14].

The basic methodology of the University of Florida Aeronomy Group involves two stages of analysis. The first stage includes (1) the use of generalized oscillator strengths obtained from electron-impact and photoabsorption data to arrive at cross sections in the high energy region, (2) a reasonable prescription for distorting the generalized oscillator strengths to provide estimates of low energy cross sections, and (3) representation of the results into more convenient semiempirical cross sections designed for the purposes of deposition calculations. In the second stage, where applications are the goal, the approach follows the flow chart of Figure VI-1, as previously described, culminating in predicted spectral intensities. All through the implementation of the method, simple analytic functions are used to input the essential data; namely, the cross sections and the self-consistently produced loss function.

Cross sections of two distinct types are considered, the type being determined not by the process of excitation so much as by the way the

cross section is best represented. All cross sections that can be characterized by a discrete, or at least a fairly limited, range of energy loss are parameterized by the functional form

$$\sigma_j(E) = \frac{q_o F}{W_j^2}\left(\frac{W_j}{E}\right)^{\Omega} \cdot \phi\left(\frac{W_j}{E}\right) \tag{1}$$

in which $\phi(W_j/E)$ is a function having either one of the two forms

$$\phi_a(W_j/E) = (1 - W_j/E)^{\nu} \tag{2a}$$

or

$$\phi_b(W_j/E) = 1 - (W_j/E)^{\tau} \quad . \tag{2b}$$

F is the strength parameter, W_j is the effective threshold for the state, q_o is the constant 6.514×10^{-14} eV^2 cm^2, and Ω, as well as ν or τ, are fitting parameters. The value of Ω determines the asymptotic E dependence and is fixed at 0.75 for all allowed states, consistent with the Born approximation. Forbidden states, in general, have values of Ω near 1.0 or 3.0, depending on the degree of forbiddenness. The strength F can be related to the optical oscillator strength for allowed states. For the case of a Rydberg series, the strength F can be stated conveniently as $F = F^*/(n - \delta)^3$ for all n, and at some convenient point, the remaining values of n to infinity can be lumped into a single effective summed cross section. Processes that can be described well by a function such as equation (1) include all ordinary electronic states, as well as most dissociative excitation and ionization states. When the state is extensively spread out, so that the average energy loss is somewhat higher than W_j, it is necessary to define a parameter $\overline{W}_j$ which gives this real average loss.

The other type of cross section is one in which the energy loss is a continuum. All usual ionization and some dissociative ionization processes must be considered in this way. Here, equation (1) is replaced by a differential cross section of the parameterized form

$$S_i(E,W) = \frac{\overline{q}_o A_o}{W^2}\left(\frac{I_i}{W}\right)^{p}\left(\frac{W}{E}\right)^{\Omega} \cdot \Phi\left(\frac{W}{E}\right) \quad , \tag{3}$$

where the extra parameter p serves to regulate the W dependence. It should be recognized that $S_i(E,W)$ is the same as $S_i(E,T)$ discussed earlier since the loss W is just the secondary energy T plus the ionization

potential I_i. Integrating over all available W produces the total cross section.

The gases of prime interest in the earth's upper atmosphere are N_2, O_2, and O. In Figures VI-2 and VI-3 the energy levels of N_2, N_2^+, O_2, and O_2^+ that have been included in the analysis are illustrated. The energy levels for atomic oxygen have been given by Green and Barth [VI-8], and since no additions have been made, it is not reproduced. Although most of the cross sections used in our analysis are the same as those of recent calculations performed in our group [VI-11, VI-15, VI-16], some additions and alterations have been made on the basis of more recent data.

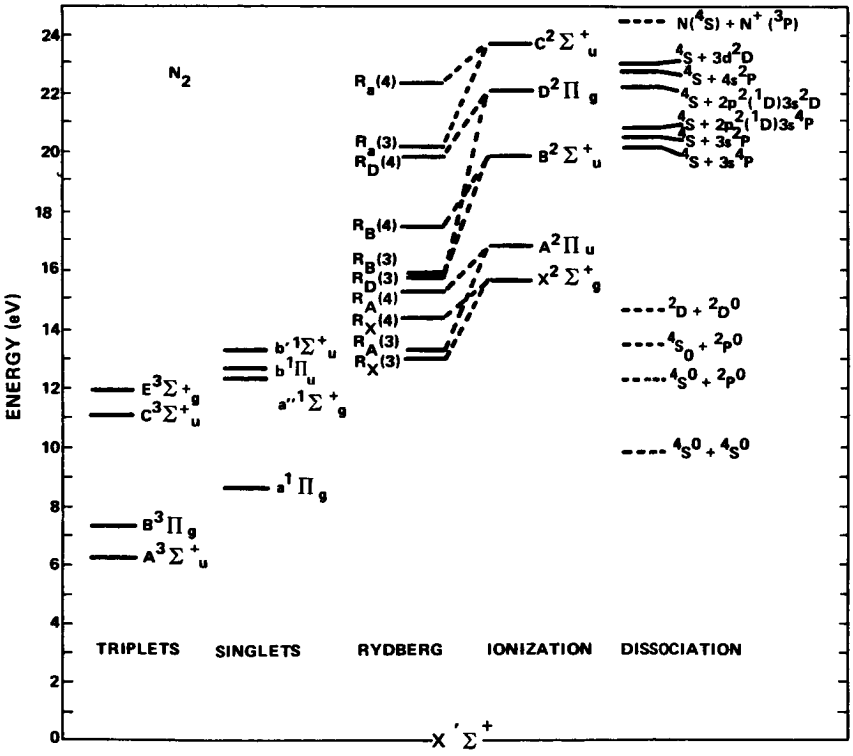


Figure VI-2. Energy level diagram for important states in N_2 and N_2^+ (dissociation levels that are drawn as dashed lines are included in the total cross section but are not treated individually; above the arbitrary cutoff at 25 eV, many additional dissociative ionization levels are known).

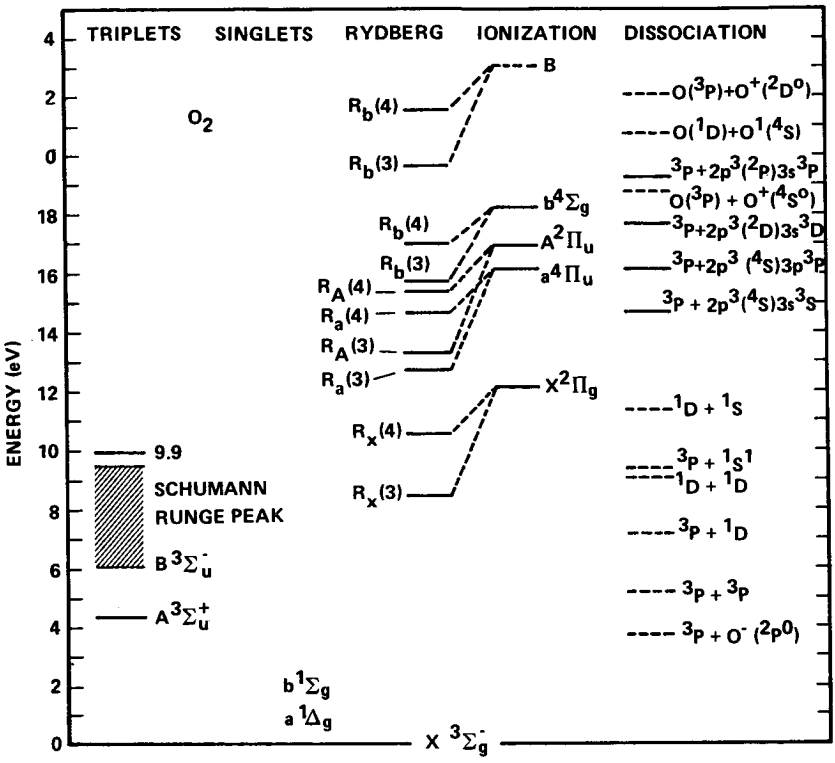


Figure VI-3. Energy level diagram for important states in O_2 and O_2^+ (dissociation levels that are drawn as dashed lines are included in the total cross section but are not treated individually; above the arbitrary cutoff at 25 eV, many additional dissociative ionization levels are known).

Electron-impact spectra for N_2 are especially abundant in the literature at both high and low impact energies. A recent review by Lassettre [VI-17] includes spectra from sources too numerous to mention individually. Since then, additional spectra have been published at both high and low energies [VI-18-VI-20]. These spectra provide the means of obtaining a rather complete set of cross sections for the many excited states, broken down into vibrational components if needed. Very noticeable in the low energy data are the singlet-triplet transitions $A^3\Sigma_u^+$, $B^3\pi_g$, $C^3\pi_u$, and to some extent $E^3\Sigma_g^+$. In addition, at low energies the very strong $a^1\pi_g$ (LBH) and the weak $a''^1\Sigma_g^+$ states are seen. Since the forbidden states have been analyzed in detail by Brinkman and Trajmar [VI-20], we have adjusted our cross sections to be reasonably compatible with theirs. The N_2 forbidden cross sections are shown in Figure VI-4. The adjusted $a^1\pi_g$ cross section agrees with our previous choice at high

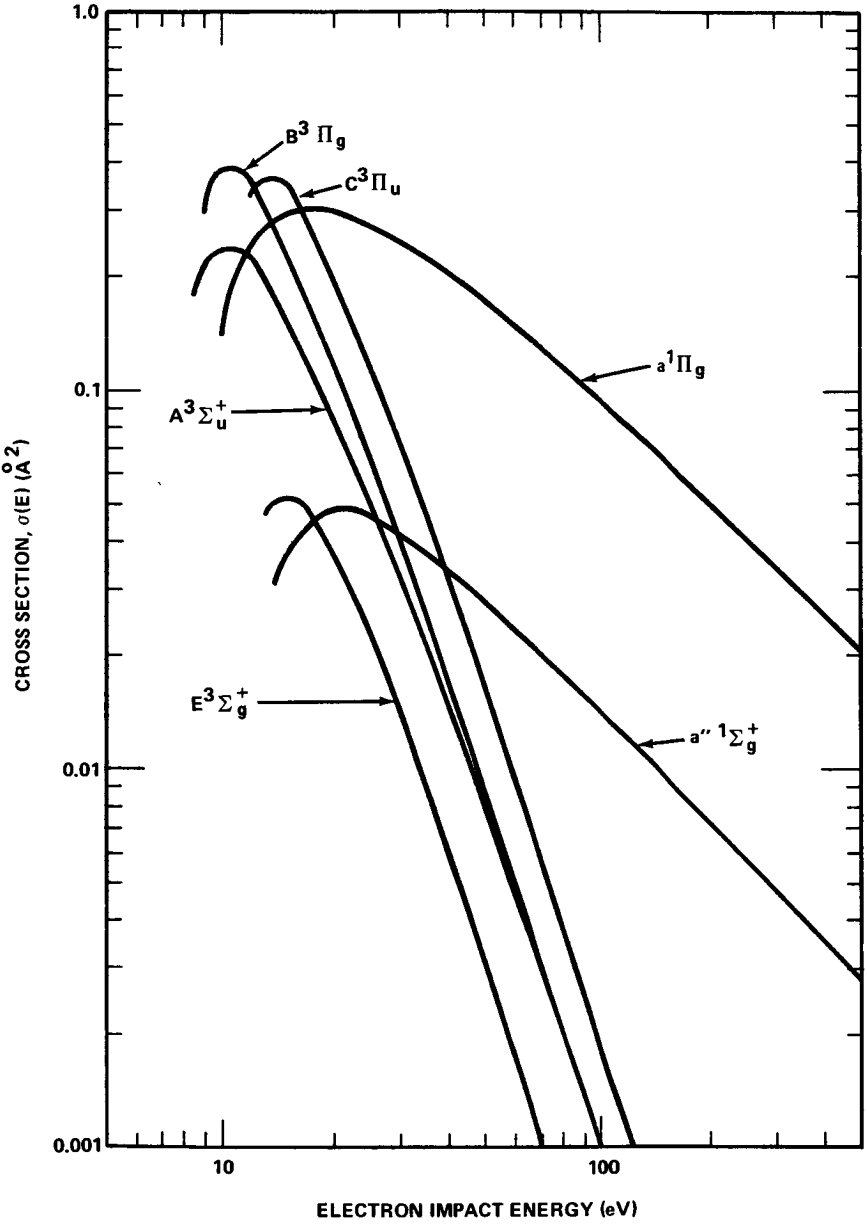


Figure VI-4. Cross sections assumed for the optically forbidden excitations in N_2.

energy and has the same asymptotic slope, but it peaks somewhat earlier and higher.

The intermediate loss region of 12.5 to 15.0 eV is resolved rather explicitly in the data of Geiger and Schröder [VI-18] and would be expected to lead to reliable relative cross sections for the allowed states $b^1\pi_u$, $b'^1\Sigma_u^+$, $R_X(3)$, and to some extent $R_A(3)$. Based on this spectrum one can estimate strengths relative to the 12.85 peak composed of five members ($v' = 0$-5) of the $b^1\pi_u$ and the first bands of the Worley-Jenkins series $R_X(3)$. Some estimates scaled to the 12.85 cross section given by Stolarski, et al. [VI-15] are shown in Figure VI-5. It is difficult to make a direct comparison with our earlier cross sections since they were based on low resolution spectra, and thus the identifications are less explicit. A very large amount of additional data on cross sections by methods other than energy-loss measurements are also available, but these are not singled out in the present work. A complete self-consistent

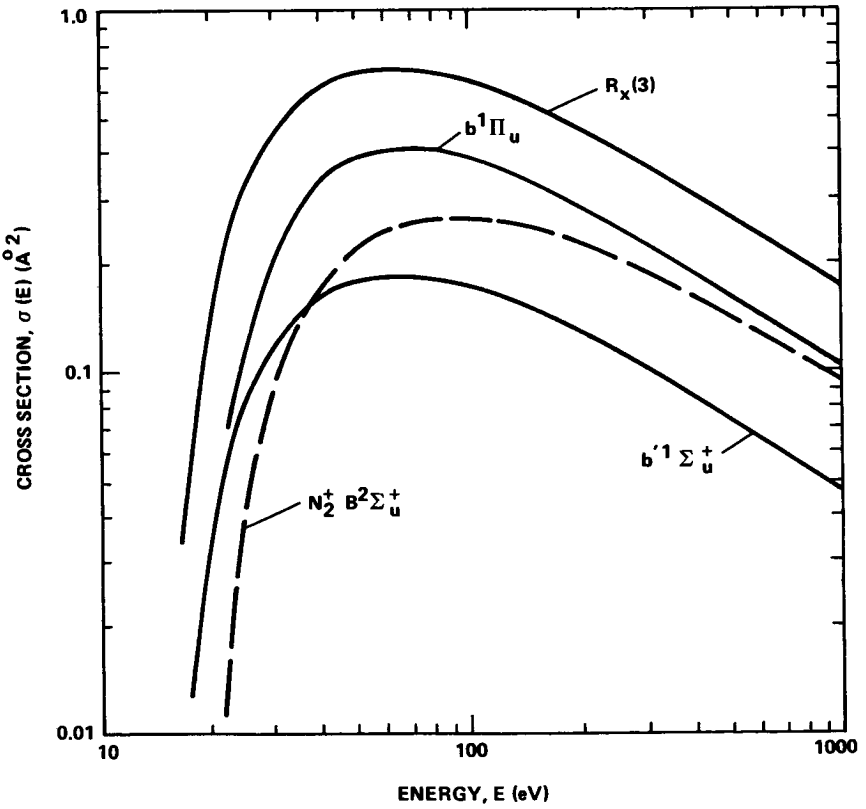


Figure VI-5. Cross sections assumed for the optically allowed excitations in N_2 and in the B state of N_2^+ (the effective cross section for 3914-Å emission is two-thirds of the plotted cross section).

reworking of all the cross sections is in order, and such an analysis is underway.

In connection with our concentration on the electron-impact excitation of $O_2(a^1\Delta_g)$, having a threshold of 0.98 eV, we must, however, include details of all competing cross sections in this neighborhood. The main competitors are $O_2(b^1\Sigma_g^+)$, $OI(^1D)$, $OI(^1S)$, and vibrational excitations in the ground state of N_2. In addition, for auroral calculations, one should include electron-electron collisions as a competitor below a few electron volts. Among these, the cross sections for $OI(^1D)$, $OI(^1S)$, and N_2 ground-state vibrations are approximately known, as is the electron-electron cross section [VI-21-VI-23]. On the other hand, O_2 $(^1\Delta_g)$ and $O_2(^1\Sigma_g^+)$ are difficult to measure directly and thus are poorly known. A reasonable upper limit on the $O_2(^1\Delta_g)$ cross section is suggested by the data of Hake and Phelps [VI-24]. Although it is not clear from their data what portion is due to $^1\Delta_g$, a maximum near 8×10^{-18} cm^2 is possible, with a width at half maximum of about 1 eV. Schultz and Dowell [VI-25] find a cross section more than two orders of magnitude less than this. In contrast, Kuppermann[2] has reported a cross section strength near 10^{-17} cm^2.

The revised shape of the cross sections for $^1\Delta_g$ and $^1\Sigma_g^+$ is markedly different from that used in earlier work [VI-6, VI-7, VI-11]. The new assumed shape is very sharply peaked, consistent with the highly forbidden nature of the transitions. It is also assumed that the cross section for $^1\Sigma_g^+$ has the same shape as $^1\Delta_g$ but peaks at a value of one-third that of $^1\Delta_g$. All these low energy cross sections, as we have taken them, are shown in Figure VI-6. Also, for completeness, the parameters of all the fits given in Figures VI-4, VI-5, and VI-6 are listed in Table VI-1.

Finally, as a standard of reference, the cross section for the $N_2^+(B^2\Sigma_u^+)$ state producing the 3914-Å (0, 0) band is taken from the data of Borst and Zipf [VI-26]. The explicit 3914-Å cross section is two-thirds of the $N_2^+(B^2\Sigma_u^+)$ cross section shown in Figure VI-5. All other ionization cross sections remain unchanged from our previous work. The total sum of ionization states plus dissociative ionization for each gas does fit rather well the data shown by Kieffer and Dunn [VI-27]. However, it is probable that the relative strengths of some of the states will have to be shifted when more information becomes available. Ajello [VI-28], for example, finds little evidence for the excitation of the $N_2^+(C^2\Sigma_u^+)$

2. Private communication, 1971.

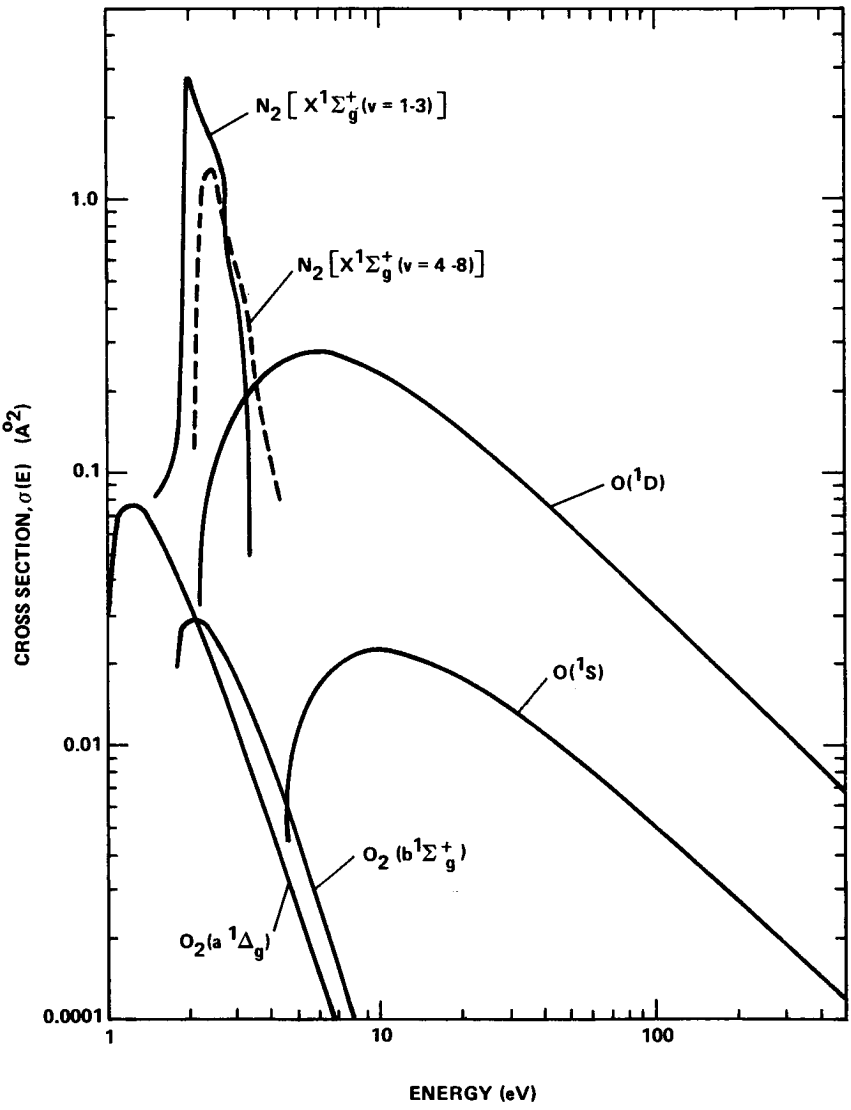


Figure VI-6. Cross sections assumed for the low lying states in N_2, O_2, and O.

state, and there is disagreement among several authors on the shape and strength of the $N_2^+(A^2\pi_u)$ Meinel cross section.

TABLE VI-1. PARAMETERS FOR ALL CROSS SECTIONS SHOWN IN FIGURES VI-4 THROUGH VI-6 (THE N_2 VIBRATIONAL COMPOSITE CROSS SECTIONS IN FIGURE VI-6 ARE FITTED TO ANALYTIC FORMS, BUT IN OUR CALCULATION OF L(E) THE CONTRIBUTIONS FROM EACH V LEVEL WERE CONSIDERED NUMERICALLY)

Species	State	W_j [a]	F	Ω	τ	ν
N_2	$A\ ^3\Sigma_u^+$	8.0 (7.4)	0.226	3.0	1.0	-
	$B\ ^3\pi_g$	8.5 (8.0)	0.178	3.0	3.0	-
	$C\ ^3\pi_u$	11.05	0.28	3.0	3.0	-
	$E\ ^3\Sigma_g^+$	11.9	0.048	3.0	3.0	-
	$a\ ^1\pi_g$	8.6 (9.1)	0.136	1.0	1.0	-
	$a''\ ^1\Sigma_g^+$	12.25	0.027	1.0	2.3	-
	$b\ ^1\pi_u$	12.5 (12.8)	0.67	0.75	-	3
	$b'\ ^1\Sigma_u^+$	13.3 (14.0)	0.33	0.75	-	3
	$R_x(3)$	13.0	1.18	0.75	-	3
	vib (1-3)	1.85	0.273	7.0	1.0	-
	vib (4-8)	2.15	0.241	9.0	1.0	-
O_2	$a\ ^1\Delta_g$	0.98	0.0005	3.0	3.0	-
	$b\ ^1\Sigma_g^+$	1.64	0.0005	3.0	3.0	-
O	^1D	1.85 (1.96)	0.010	1.0	-	2
	^1S	4.18	0.0042	1.0	0.5	-
N_2^+	$B\ ^2\Sigma_u^+$	I=18.75	A_o=0.368	Ω=0.94	τ=0.152	

a. When the average energy loss is significantly different from W_j it is given in parentheses.

E. Dissociative Excitation

One of the more significant attempts at updating that we have done relates to the dissociative excitation and simultaneous dissociative excitation and ionization process, whereby atomic states, neutral or ionic, are excited by electron impact on the parent diatomic molecule. This was prompted by the discovery that dissociative excitation may be the major process in the excitation of many significant atmospheric emissions; e.g., the 1200-Å radiation from NI.

Production of excited atomic species by electron impact on molecules involves excitation of the molecule to the dissociative part of either the neutral or ionic molecular potential-energy curves. The ideas involved are treated in some detail by Herzberg [VI-29]. Three possibilities arise. In the first case, there is the direct excitation from the ground molecular state to a repulsive state or to a bound state above its dissociation limit. The second and third possibilities relate to two-step processes involving excitation to an intermediate bound state followed either by predissociation or by cascade to a lower repulsive state. Often, two or even all three possibilities may be involved in the excitation of an atomic state. This makes it difficult to decide whether the transition is optically allowed, is optically forbidden, or involves a change of multiplicity.

When an electronic transition to a repulsive state occurs, the impending dissociation usually takes place in an amount of time that is small in comparison to that for rotation. The dissociation products will then be travelling in a direction that is approximately characteristic of the vibratory motion. Then, provided the excitation probability depends on the relative orientation of the exciting beam and any preferred axis of the molecule, the dissociation products will have corresponding angular asymmetry [VI-30, VI-31].

In an experiment, the production of an excited atomic species is detected by monitoring the radiation from it. This tells us nothing about the state of excitation of the other fragment, although some information can be obtained from the study of appearance potentials. Furthermore, very little information about the intermediate repulsive molecular state involved in the process is obtained from the cross section measurements. A study of the angular distributions, combined with the above symmetry arguments, can be helpful in determining the excited molecular state, and some studies along these lines have been made for H_2 [VI-32].

Considerable cross-sectional data are available from radiation following electron-impact induced dissociation processes in N_2 and O_2. A number of independent measurements exist for some cross sections, and except for the data of Sroka [VI-33], are reasonably consistent with each other. The cross sections of Sroka are on the average a factor of about six larger than those of Mumma, Ajello, and Aarts, et al.

[VI-28, VI-34-VI-36]. No satisfactory explanation is immediately available, and in presenting our estimates, his values have all been reduced by this factor of six. The semiempirical parameters for the cross sections in the format of equation (1), as fitted to an average of the available data, are presented in Table VI-2. In addition, the functions generated from these parameters are given in Figures VI-7 through VI-9 for NI and NII, and also OI and OII.

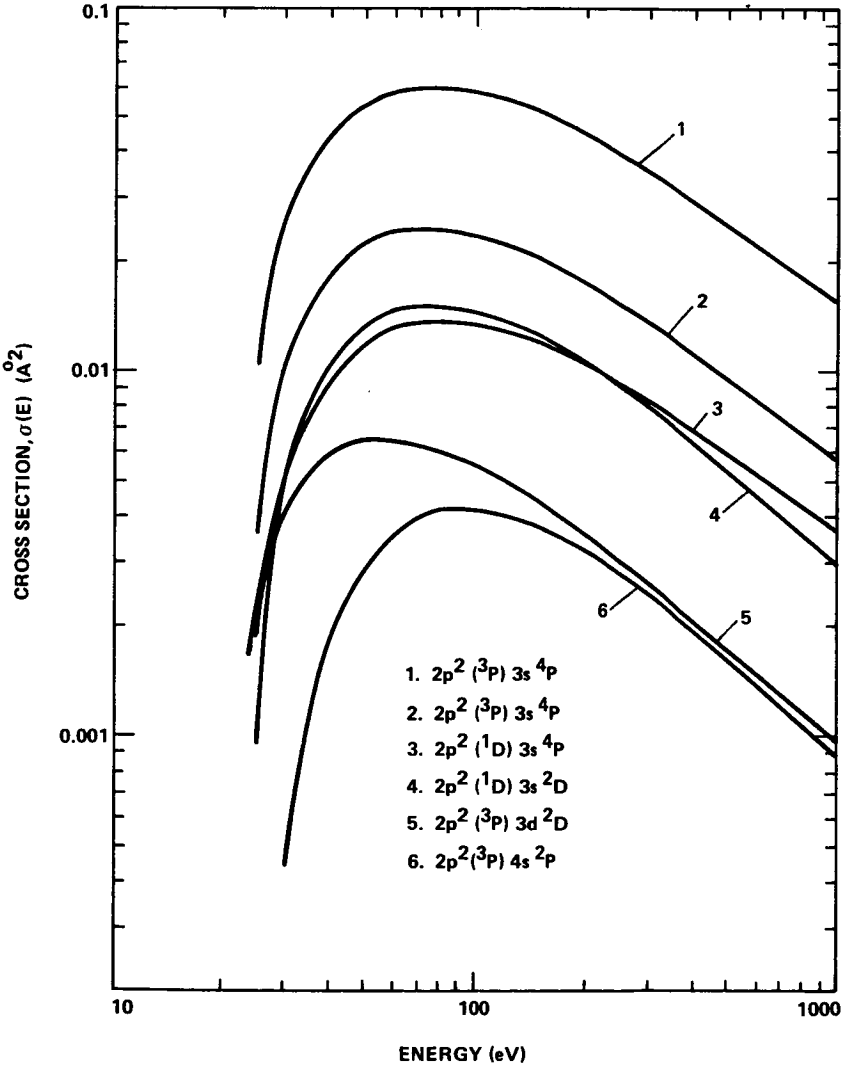


Figure VI-7. Cross sections for dissociative excitations in N_2 leading to excited fragments.

TABLE VI-2. PARAMETERS FOR THE DISSOCIATIVE EXCITATION AND IONIZATION CROSS SECTIONS SHOWN IN FIGURES VI-7 THROUGH VI-9 (ALL PARAMETERS ARE CONSISTENT WITH EQUATION (1), UNDER THE ASSUMPTION THAT $\bar{W}_j$ IS 10 PERCENT LARGER THAN THE W_j GIVEN IN THE THIRD COLUMN OF THE TABLE; PROMINENT EMISSIONS FROM THE EXCITED FRAGMENTS ARE IDENTIFIED IN PARENTHESES ALONG WITH THE STATE)

Species	State	W_j(eV)	F	Ω	τ	ν
N_2	$2p^2(^3P)\,3s\ ^4P(1200)$	21.0	0.20	0.75	-	2
	$2p^2(^1D)\,3s\ ^3P(1134)$	21.0	0.047	0.75	-	2
	$2p^2(^3P)\,3s\ ^2P(1493)$	21.0	0.085	0.80	-	2
	$2p^2(^1D)\,3s\ ^2D(1243)$	22.0	0.060	0.85	-	2
	$2p^2(^3P)\,3d\ ^2D(1164)$	23.0	0.043	1.00	0.4	-
	$2p^2(^3P)\,4s\ ^2P(1177)$	23.0	0.031	1.00	-	3
N_2^+	$2s\ 2p^3\ 3D\ (1084)$	36.0	0.94	1.00	0.4	-
	$2s\ 2p^3\ ^3P\ (916)$	38.0	0.33	1.00	-	2
	$2s\ 2p^3\ ^1D\ (776)$	42.0	0.074	0.75	-	2
	$2p(^2P)\,3s\ ^1P\ (747)$	43.0	0.13	1.00	-	2
	$2p(^2P)\,3s\ ^3P\ (672)$	43.0	0.13	1.00	-	2
	$2p(^2P)\,3d\ ^3F\ (5003)$	48.0	0.086	1.00	-	2
	$2p(^2P)\,3d\ ^3D\ (533)$	70.0	0.17	1.00	-	2
O_2	$2p^3(^4S)\,3s\ ^3S\ (1304)$	14.6	0.085	0.75	-	3
	$2p^3(^4S)\,3p\ ^3P\ (8447)$	16.1	0.067	0.90	-	3
	$2p^3(^2D)\,3s\ ^3D\ (990)$	17.6	0.032	1.00	-	3
	$2p^3(^2P)\,3s\ ^3P\ (879)$	19.2	0.016	0.75	-	3
O_2^+	$2s\ 2p^4\ ^4P\ (833)$	34.0	0.26	1.00	-	3
	$2s\ 2p^4\ ^2D\ (717)$	55.0	0.17	1.00	-	2

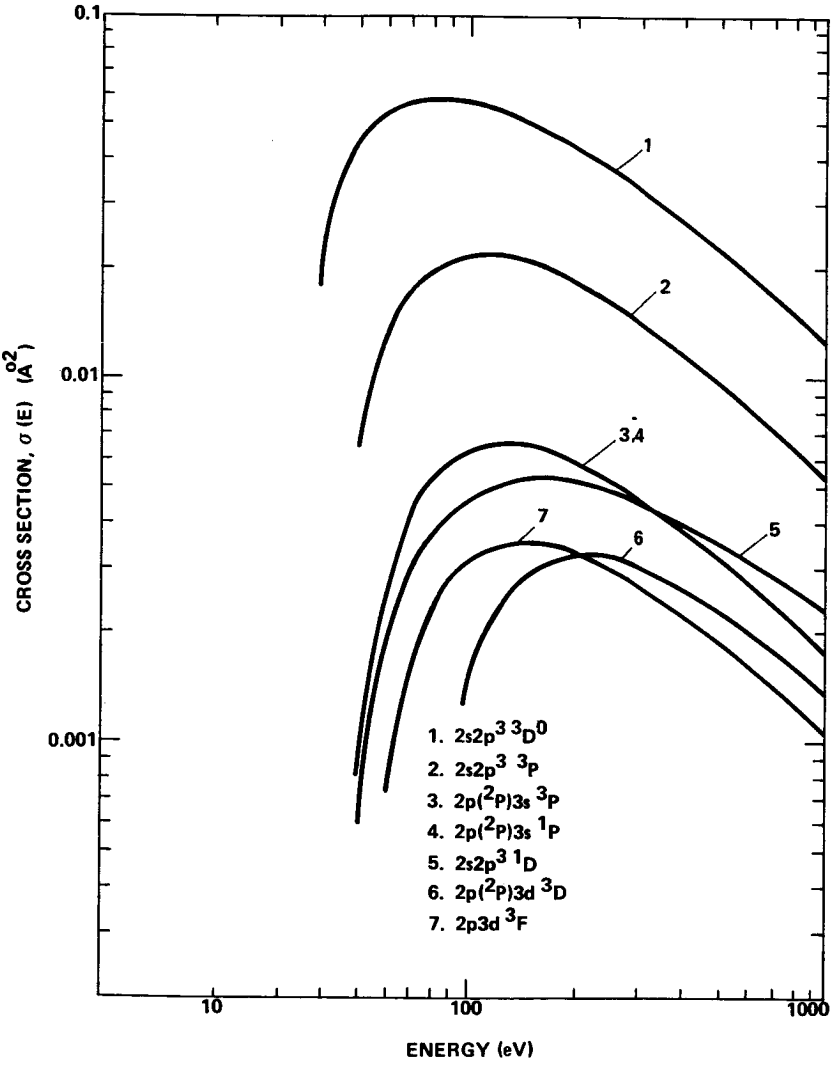


Figure VI-8. Cross sections for dissociative ionizations in N_2 leading to excited fragments.

In constructing loss functions from these cross sections, it should be remembered that the dissociation fragments often come out with some kinetic energy. This is beautifully revealed in the experiments of Dunn and Kieffer [VI-37] on dissociative ionization of H_2, where the H^+ ion comes out with an average kinetic energy of 6 to 8 eV. To allow approximately for this kinetic energy, the energy loss in any dissociative excitation process has been taken to be the appearance potential plus 10 percent.

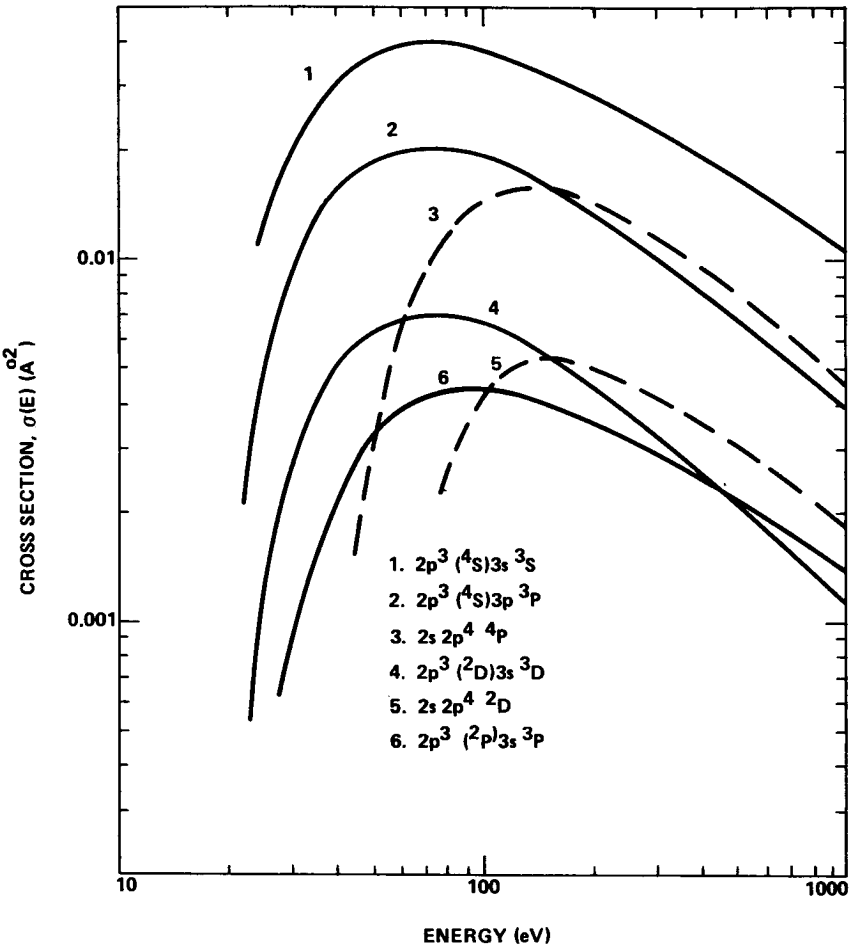


Figure VI-9. Cross sections for dissociative excitations and ionizations in O_2 leading to excited fragments (both the dashed curves involve ionizations).

F. Thermal Excitation

As discussed earlier, a part of the primary energy goes into the ambient electron gas via elastic electron-electron collisions. Auroral activity, therefore, raises the ambient electron temperature [VI-38]. For low lying states, significant excitation is also possible by the energetic electrons in the Maxwellian tail of the energy distribution [VI-3]. Assuming this distribution, the excitation rate $\underset{\sim}{k}$ is related to the cross section $\sigma(E)$ by

$$\underset{\sim}{k} = \left(\frac{1}{\pi\mu}\right)^{1/2} \left(\frac{2}{kT_e}\right)^{3/2} \int_0^\infty \sigma(E)\, E \exp\left(-\frac{E}{kT_e}\right) dE \quad , \qquad (4)$$

where T_e is the electron temperature, μ is the reduced mass, and k is the Boltzmann constant. Obviously, the excitation rate is a sensitive function of the electron temperature.

A simple practical way of utilizing this relationship is to express the relevant portion of the appropriate cross section in some convenient analytic form [VI-1]. In the process of fitting, care should be exercised to reproduce accurately the portion near the threshold. The semiempirical representations of cross sections that we have adopted for the energy degradation calculations are convenient for use in equation (4), since they lead to exponential integrals. It must be remembered, however, that in most cases our representations are primarily accurate for the region encompassing the peak and continuing out to high energy.

In the present work, excitations to the $O_2(a^1\Delta_g)$ state can be significantly enhanced by the heated electrons. We have therefore calculated the temperature dependence of the thermal excitation rate for this state using the cross section proposed in section D. The results are shown in Figure VI-10. In this particular case, because of the highly forbidden nature of the transition with a peak just above threshold, our analytic representation would be expected to be satisfactory.

It is interesting to note that the excitation rate is intimately related to the heated electron cooling rate via that particular inelastic excitation. It is very obvious that the cooling rate equals the excitation threshold energy times the excitation energy. In fact, the excitation of very low lying states, including the rotational and vibrational states as well as the ground-state excitations in atomic oxygen, is a very efficient cooling mechanism for the hot electron gas [VI-39].

G. Efficiencies

The methods described previously have been applied to the question of energy deposition resulting from complete degradation of an electron having any energy up to 1 keV locally absorbed at 120 km. For this purpose we use the number density ratios for O, O_2, and N_2 of 0.138, 0.136, and 0.726 taken from the CIRA model atmosphere [VI-40]. In all, approximately 200 cross sections and corresponding energy losses assumed for the states of these gases are involved in the total effective loss function presented in Figure VI-11. For calculation of the populations $J_j(E)$ and of the intermediate electron spectra $n(E,T)$, the loss function $L(E)$

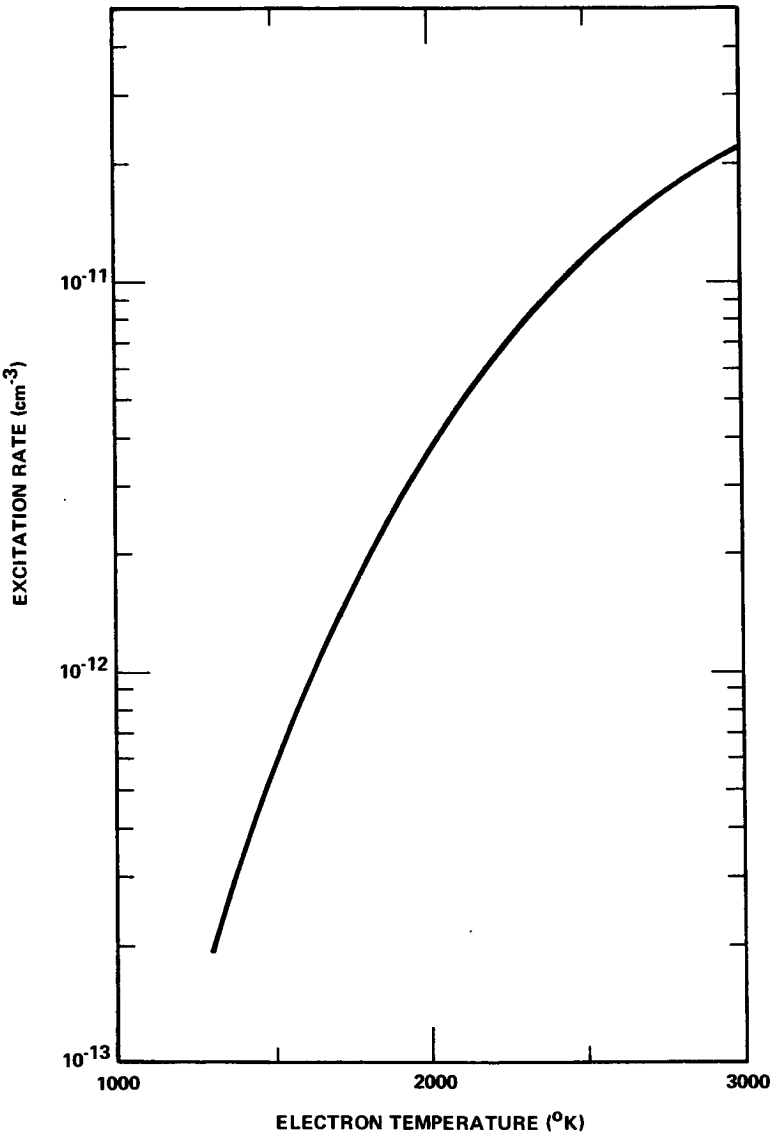


Figure VI-10. Rate for thermal excitation of $O_2(a^1\Delta_g)$ as a function of electron temperature.

is represented by an inverse power series, which provides an accurate fit for energies above 20 eV. Here,

$$L(E) = \left[\frac{q_o Z^+}{2R_e}\left(\frac{E}{A}\right)^{\Omega} + \left(\frac{E}{B}\right)^{M} + \left(\frac{E}{C}\right)^{\Lambda}\right]^{-1}, \qquad (5)$$

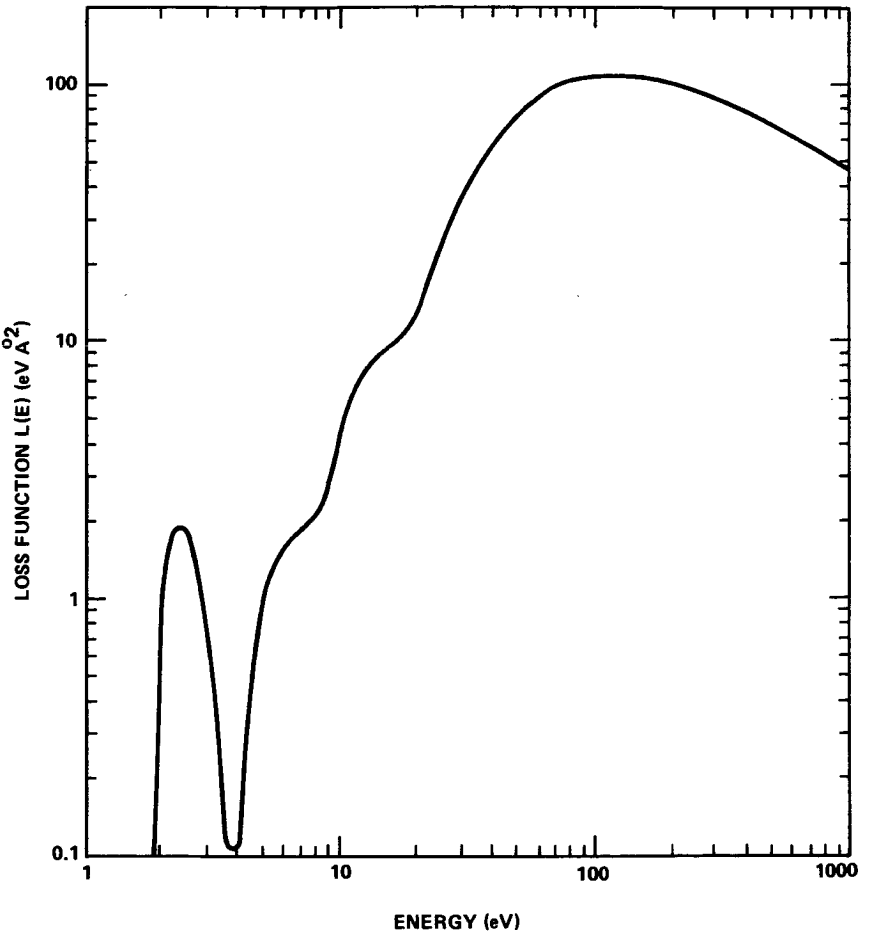


Figure VI-11. Composite loss function for an O, O_2, N_2 mixture having the fractional composition 0.138, 0.136, and 0.726, respectively (120 km).

where q_o is the same constant seen in equation (1), R_e is the Rydberg energy, and where the other parameters have the following least squares values: $Z^+ = 10.14$, $\Omega = 0.70$, $M = -0.634$, $\Lambda = -3.50$, $A = 100.0$, $B = 154.5$, and $C = 44.3$. For the small portions of the calculation below 20 eV, numerical methods are used.

The relative populations are presented in Table VI-3 for several states following degradation of a 1-keV electron. It is also instructive to give the results as efficiencies; i.e., the fractional amount of primary energy ending up in a particular state. This is found by dividing the product $\overline{W}_j J(E)$ by the initial kinetic energy E of the incident electron.

TABLE VI-3. POPULATIONS OF LOW LYING STATES IN N_2, O_2, AND O FOLLOWING COMPLETE DEGRADATION OF A 0.1- AND A 1.0-KEV ELECTRON AT 120-KM (AN AMBIENT ELECTRON DENSITY OF 10^6 CM^{-3} AT 2000 °K IS ASSUMED)

State	Starting energy	
	0.1 keV	1.0 keV
N_2 vib (v'=1-3)	6.2	54.0
N_2 vib (v'=4-8)	2.7	24.0
$O_2(a^1\Delta_g)$	0.41	3.5
$O_2(b^1\Sigma_g^+)$	0.073	0.64
$O(^1D)$	2.5	22.0
$O(^1S)$	0.10	0.90
N_2^+(3914 Å)	0.11	1.3

The results for a number of excited states and also for the dissociative states are given in Figures VI-12 and VI-13, both for the case of an electron density of 10^6 cm^{-3} at 2000 °K and for a density of zero.

Several features are immediately apparent. The efficiency of the low lying $^1\Delta_g$ state is, as expected, significantly affected by the presence of the hot ambient electron gas, whereas the others are not. From a few hundred electron volts upward in energy, the efficiencies of all states approach constants. However, the behavior at lower energy depends on whether the state is forbidden or allowed. Efficiencies for allowed transitions start off slowly and then increase steadily to their limiting value as the energy increases. Forbidden transitions, on the other hand, go rapidly through a rather pronounced peak at low energy, consistent with the sharpness of the cross section peak itself, and then settle down to roughly 10 to 20 percent of the peak efficiency. This rather graphically illustrates that a forbidden state with a low threshold can be excited very strongly compared to the 3914-Å emission if the incident electron spectrum is strongly biased toward low energy particles.

A calculation of electron volts per ion pair was also done by summing over populations of all ionization levels. Included was the assumption that on average, Rydberg levels decay by means of autoionization 50 percent of the time. This calculation leads to a value of 36.1 for the electron volt per ion pair at 1.0 keV.

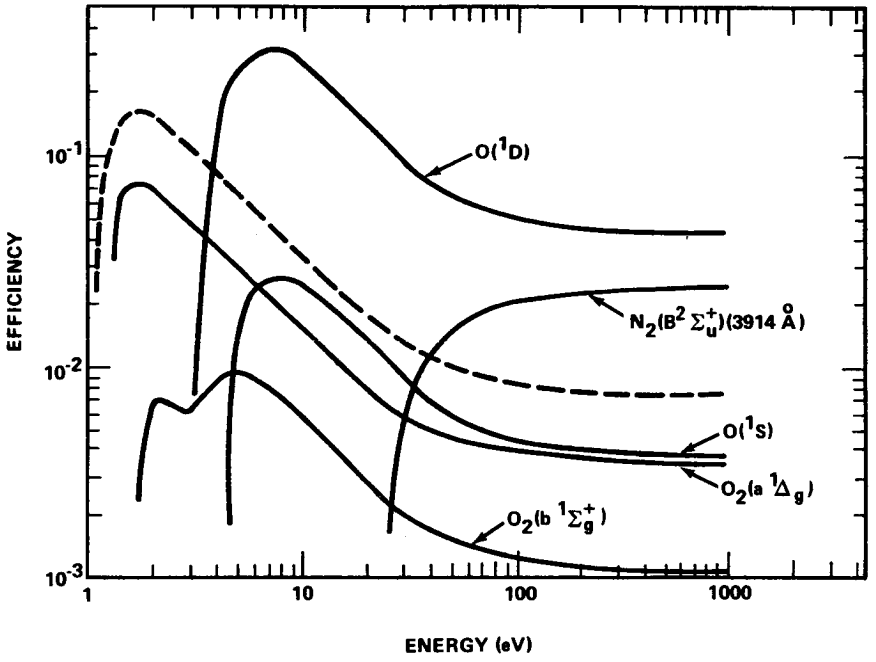


Figure VI-12. Efficiencies of the low lying states in O, O_2, and N_2 corresponding to local degradation at 120 km (an electron density of 10^6 at a temperature of 2000 °K is assumed; the dashed curve shows the effect on $^1\Delta_g$ when the electron density is zero; the effect on the remaining states is very small).

As a final comment, the total efficiency of all processes we have considered is about 0.95. This implies that roughly 50 eV of an incident 1.0-keV electron are available for such processes as heating the electron gas.

H. The Problem of $O_2(a^1\Delta_g)$

The strong enhancement of the emissions from the $O_2(a^1\Delta_g)$ and $O_2(b^1\Sigma_g^+)$ states, both in airglow and aurora, has received considerable attention recently and has been reviewed in the literature [VI-12-VI-14]. The situation for $^1\Delta_g$ in dayglow is reasonably well established as being dominated by ozone photolysis in the Hartley continuum [VI-41]. The $^1\Sigma_g^+$ dayglow profile similarly suggests a large contribution from ozone photolysis in the 60-km region but it also indicates resonance fluorescence and energy transfer from O(^{1}D) at higher altitude [VI-42].

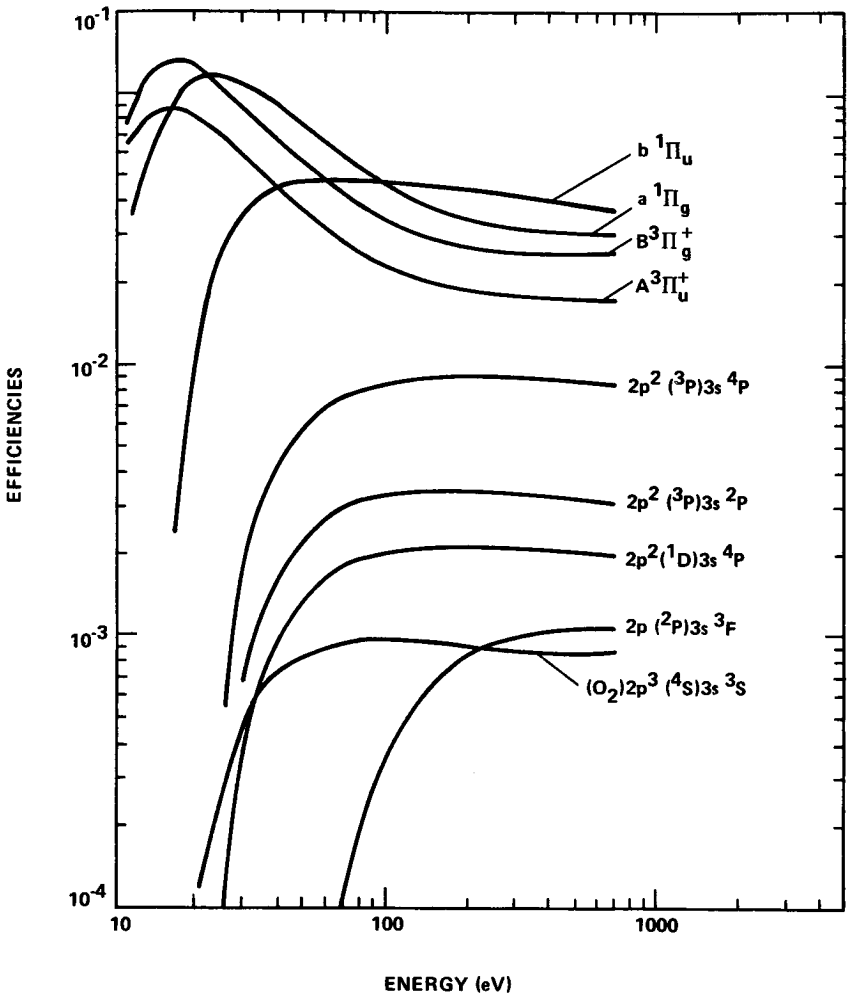


Figure VI-13. Efficiencies of representative electronic and dissociative levels in O_2 and N_2.

In nightglow and aurora, however, the explanation is not yet clear. In fact, the very large intensities for $^1\Delta_g$ emission at 1.27 μ compared to the 3914-Å (0, 0) band in aurora lead one to serious problems with the energy balance [VI-13, VI-14, VI-43]. As pointed out by Noxon [VI-13], some auroral enhancement measurements of the intensity ratio I(1.27 μ)/I(3914 Å) suggest that the energy going into the $^1\Delta_g$ state is comparable to the total energy in the incident electron flux. If one ignores the spatial aspects of the data, ratios of the order of a few hundred must

be explained. The spatial aspects produce apparent instantaneous ratios much more extreme than these, and involve a more sophisticated interpretation.

Despite many uncertainties it appears that only a small fraction of the $^1\Delta_g$ excitation can come from electron impact. Let us now investigate alternate processes. Consider first the redistribution of energy going originally into other channels. A significant fraction of the primary energy goes into excitation of Rydberg or non-Rydberg excited states of the various atmospheric species. Some of the excited states in O_2 can, of course, populate the $^1\Delta_g$ state by cascading, but this contribution cannot be important. There is, on the other hand, an indirect way in which the excited states can contribute. Some of the ultraviolet emissions from these states lie in the Schumann-Runge and the Hartley continua and can be absorbed by the molecular oxygen and ozone in the regions below the aurora. The relevant photochemistry has recently been studied by Isaksen et al. [VI-44] with the result that only a feeble emission of the infrared atmospheric band system can be expected. In their studies the possibility of some of the Rydberg states cascading to the upper state of the LBH system, for example, was not taken into account. Their estimates should be regarded as upper limits, and the conclusion that the ultraviolet absorption is a negligible contributor remains unaltered.

Also of interest is the role played by the low lying metastable $N(^2D)$ and $O(^1D)$ atoms. Although the reaction of $N(^2D)$ with N_2 is spin-forbidden, reaction with O_2 is possible, and excitation of $^1\Delta_g$ can occur by means of

$$N(^2D) + O_2 \rightarrow N + O_2(^1\Delta_g \text{ or } ^1\Sigma_g^+) \quad . \tag{6}$$

There is, however, severe competition from the alternate reaction

$$N(^2D) + O_2 \rightarrow NO + O \quad . \tag{7}$$

Experimental data of Lin and Kaufman [VI-45] and of Black, et al. [VI-46] suggest that reaction (7) may be very fast. This would then mean that $O_2(^1\Delta_g)$ production via $N(^2D)$ is not significant. It appears that $O(^1D)$ reacting directly with O_2 is also ineffective. At the altitudes of interest, $O(^1D)$ is deactivated by N_2 and O_2, with quenching coefficients of 5 to 10×10^{-11} and 4 to 6×10^{-11}, respectively. Keeping in mind the relative abundances of N_2 and O_2, perhaps only 10 to 20 percent of the effective collisions are with O_2. Furthermore, although the outcome of quenching by O_2 is still controversial [VI-47, VI-48], it appears that $O_2(^1\Delta_g)$ is produced with only small probability.

It may be possible to recover some of the $O(^1D)$ energy that goes into vibrational excitation of N_2. In this connection, it may be noted that N_2 vibrational excitation is also efficiently produced by electron impact, as is apparent from the results of section G, in which a 1-keV electron causes vibrational excitations having $v \geq 4$. In addition, there is the possibility of excitation through the vibrational interchange process [VI-10].

$$N_2(A^3\Sigma_u^+, v) + N_2(X^1\Sigma_g^+, v=0) \rightarrow N_2(A^3\Sigma_u^+, v') + N_2(X^1\Sigma_g^+, v'') \quad (8)$$

involving the metastable upper state of the VK system (and presumably involving $a^1\pi_g$ also). The net result may be a highly effective N_2 vibrational temperature. It is now energetically possible for the following reaction to go:

$$O_2 + N_2^*(v' \geq 4) \rightarrow O_2(^1\Delta_g) + N_2^*(V'' < V') \quad . \quad (9)$$

Detailed balancing considerations based on the inverse reaction suggest that reaction (9) should not have a large cross section. Nevertheless, assuming one out of five collisions is with O_2 and assuming one-half of the O_2 collisions excite $O_2(^1\Delta_g)$, one obtains about three excitations for each 3914-Å excitation.

Approximately one-half of the primary energy goes into the potential energy of the ions. In equilibrium, part of the energy flowing into this channel reappears as kinetic energy of the products formed in dissociative recombination of the molecular ions. Another part goes into metastable potential energy of the same products, and part takes a more complex path of charge transfer or ion-atom interchange before eventually ending up as kinetic or potential energy. Nowhere in any of these latter rearrangements of energy does there seem to be any scope for creating much $O_2(^1\Delta_g)$. Thus, this portion of the energy is also apparently wasted as far as $^1\Delta_g$ is concerned.

Let us now turn our attention to the last of the channels open for the primary energy, that part going into the heating of the ambient electron gas. It is conjectured that about 50 eV out of every 1 keV of primary energy goes into the heating. Based on our cooling rates presented in Figure VI-10 and the electron temperature in aurora as calculated by Walker and Rees [VI-38], it is found that at 120 km about four $O_2(^1\Delta_g)$ excitations are produced for each one of 3914 Å. The corresponding figure at 150 km is eight. The Walker and Rees estimate of the electron temperature may be a lower limit because they did not take into account heat sources such as vibrationally heated N_2 molecules. On the other hand, they did not include molecular O_2 cooling or cooling by excitation of

the fine-structure levels in atomic oxygen. It is difficult to estimate the net effect of these omissions, yet it seems safe to assume an excitation ratio of five.

Summing up, for every photon of 3914 Å excited in a type A aurora, our estimates of the contributions from various processes is shown in Table VI-4. It is clear from the earlier comments on the experimentally observed intensities that we have yet to search for additional sources. One possibility for enhancement lies in the kinetic energy of the dissociation fragments. It has been customary to assume this energy to be degraded into thermal motion. To a large extent it may be true inasmuch as enhanced temperatures have been reported in aurora, but the reaction

$$O_{fast} + O_2 \rightarrow O_2(^1\Delta_g) + O_{slow} \qquad (10)$$

might be fast if any curve crossing is involved. Assuming that there is indeed a curve crossing and that the cross section is about four to five times larger than that for elastic collisions with N_2 and O_2, one can estimate that about one in every two fast oxygen atoms produces a $^1\Delta_g$ excitation.

TABLE VI-4. ESTIMATED RELATIVE CONTRIBUTIONS TO $O_2(a^1\Delta_g)$ EXCITATION ALTERNATIVE MECHANISMS, REFERRED TO ONE 3914-Å PHOTON PRODUCED

Electron Impact	3
Auroral Uv	Negligible
Metastable $N(^2D)$ and $O(^1D)$	2
N_2^* $(v \geq 4)$	3
Thermal Electrons	7

a. This value can be increased by as much as an order of magnitude if an allowed shape is assumed with a peak value of 10^{-17} cm^2.

With this in mind, let us analyze the situation in an aurora occurring on March 26, 1967 at 2207 hours [VI-38]. The NO^+ density at 110 km in this aurora was about 1.8×10^6 cm^{-3}. The corresponding volume emission rate for 3914 Å was about 3×10^4 photons cm^{-3} sec^{-1}. Using the appropriate rate coefficients, the dissociative recombinations of NO^+ yield about 3.2×10^5 fast oxygen atoms. If half of these fast atoms produce an excitation of $^1\Delta_g$, then an excitation rate of 1.6×10^5

is obtained, and this is a factor of eight larger than that of 3914 Å. Some excitation of $O_2(^1\Delta_g)$ should also be possible using either of the following reactions:

$$O_2^+ + N \rightarrow NO^+ + O_{fast} \qquad (11)$$

or

$$O_2^+ + NO \rightarrow NO^+ + O_2(^1\Delta_g) \quad . \qquad (12)$$

The contribution from these reactions is not expected to be large.

One is forced to the conclusion that additional mechanisms must involve energy sources other than the primaries. Numerous experimental observations, as well as theoretical considerations, suggest that large electric fields are generated during auroral activity, and several authors have discussed the possibility of deriving energy from the field [VI-13, VI-14, VI-49]. Cole's mechanism[3] involves accelerating the ions during its mean free path and therefore can operate effectively only at the higher altitudes where the gyro-frequency is greater than the collision frequency. Other suggestions utilize the frictional heating of the auroral ionospheric electrons. It is very difficult to advance observational evidence to support this hypothesis because until now only one direct measurement of the electron temperature in an aurora has been reported [VI-50]. This measurement did show very high electron temperature in the upleg of the rocket flight, but in the downleg the temperature was close to normal. The implication is that very high electron temperatures do not exist on a regular basis during auroral activity. Detailed theoretical considerations also support this implication. Walker and Rees [VI-38] found that at lower altitude the electron heating by secondaries exceeds that by the electric field for reasonable field strengths. They note, however, that the electric field heating of the ions is much more significant [VI-51]. Because of this, Walker [VI-49] has suggested the reaction

$$O_2^+ + O_2(^3\Sigma) \rightarrow O_2(^1\Delta_g) + O_2^+ \quad . \qquad (13)$$

A large, but not physically unreasonable, cross section for the reaction was required. Difficulties arise, however, when we try to tie in the recent auroral ion composition measurements as reported by Donahue et al. [VI-52]. These measurements indicate that O_2^+ ions are insignificant in the altitude range of interest and would require a huge cross section for Walker's mechanism to be successful.

3. Private communication.

Because the electric field is a good candidate established by experimental observations, it appears that there should exist a way of exciting $O_2(^1\Delta_g)$ by means of energetic NO^+ ions. Perhaps a mechanism such as

$$NO^+ + O_2 \rightarrow (NO_3^+)^* \rightarrow NO^+ + O_2(^1\Delta_g) \qquad (14)$$

is involved, indicating the existence of a transient $(NO_3^+)^*$ complex analogous to the negative ion complexes associated with electron-impact vibrational excitation of diatomic molecules.

In summary, if our assumed atomic cross sections are approximately correct, the problem of intense 1.27-μ aurora remains unresolved. At least for the most intense events, our calculations strongly indicate a need for external energy sources. Electric fields appear to be the most obvious source, inasmuch as fields of the right magnitude have been measured. It may also be noted that, during certain auroral displays, intense heating of the neutral atmosphere takes place. The implications of this sudden heating have not been analyzed in detail as yet, but an analogous situation in stable subauroral arcs has been examined by Dickinson and Roble [VI-53]. They have shown that in response to this heating, an upward convection current is set up at the location of the heating with the concommitant large scale subsidence elsewhere. Such a circulation system in an aurora could spread molecular oxygen to higher altitudes where its excitation could be more effective. Alternatively, the atomic oxygen in the subsiding air mass may recombine in the denser atmosphere to form $O_2(^1\Delta_g)$ molecules.

I. Discussion and Conclusion

In the preceding section, we illustrated a case in which the detailed shape and magnitude of an atomic cross section has very important and direct aeronomical consequences. We found, in particular, that the ratio of the number of 1.27-μ photons to 3914-Å photons is very sensitive to the assigned $O_2(^1\Delta_g)$ cross section. In view of the large uncertainty in this cross section, we have examined the consequences of the assumption that the cross section has an allowed shape that peaks at about 10^{-17} cm^2. One then calculates an order of magnitude increase in the electron impact contribution to the ratio. While this would bring us into the range of some of the lower reported ratios (approximately 50), it fails to account for the higher ratios (approximately 1000), which is why we have sought alternate mechanisms for the 1.27-μ emission.

While in that particular instance the direct transformation of electron impact energy to photons fails to explain the observations, it must be emphasized that, for most other electronic states, such calculations do explain the observations reasonably well. Thus, it would appear that the microscopic-approach strategy of using analytic sets of cross sections in conjunction with the continuous slowdown approximation to deal with complex spectral emission phenomena has been and should continue to be fruitful. While this approach was initiated several years ago, it must be conceded that is it still in a primitive phase. We have yet to treat the breakdown of electronic states into vibrational and rotational substates, nor have we included all the associated atomic and molecular processes involved in all the atmospheric gases. However, by using systematic semiempirical characterizations of classes of input cross sections that group together fine components in an organized way, the microscopic description of the energy division is quite manageable. Indeed, it becomes essentially a matter of bookkeeping, which computers can do very effectively. Thus, with the microscopic approach, we are now able to go beyond the theory of stopping power based upon the Bethe-Born approximation in a way that contains the Bethe-Born stopping power theory at high energies.

It is interesting to note that our microscopic approach, which has been developed in response to pure aeronomical problems, may have a number of other important applications. For example, at a recent conference on the biological effects of radiation quality, several papers [VI-54-VI-56] pointed to secondary electrons as the important intermediaries in the action of the radiation. In addition, Rossi [VI-57] suggested that a cell must be damaged by two hits in proximity; e.g., within 50 Å, to make it incapable of replicating. In view of the ranges involved, one might reasonably conjecture that if a primary particle and a secondary electron or a secondary and a tertiary electron hit both strands of a DNA molecule within a small radius, then the replication mechanism would be deactivated. In this case, the normal enzyme repair mechanism could not use the second strand as a template for the repair of the first. It is too early to say whether this specific model of radiation-induced cell death is the correct one. Nevertheless, it provides a heuristic approach fostered by developments in aeronomy that may lead to an understanding of biological effects of radiation quality. The essential point is that the upper atmospheric spectroscopist, by virtue of the very detailed and comprehensive nature of his in situ spectral observations, has imposed a severe task upon the theoretical aeronomer. In response to this task, the theoretical aeronomer has developed methods for estimating reasonably accurate and complete sets of atomic properties and for inputting these properties into a detailed microscopic calculation of their effects. This same approach clearly has other important applications; e.g., the explanation of laser excitations by electron impact and of various types of gaseous discharge phenomena.

The atomic physics that one does in this type of work is quite different from the usual work carried out by an atomic physicist interested in atomic physics for its own sake. However, while dealing with the demanding requirement of aeronomy, it has also been necessary to develop simple techniques for dealing with complex atoms which, to some extent, add to our understanding of atomic physics [VI-58-VI-60].

Finally, we might mention the importance of the $O_2(^1\Delta_g)$ problem itself as discussed at a recent International Conference on Singlet Molecular Oxygen and Its Role in Environmental Sciences [VI-61]. Many of the papers pointed to the importance of this active form of oxygen in smog reactivity [VI-62]. Here again one has an example of how a problem first studied in connection with upper atmospheric [VI-63] spectral observations may lead to an understanding of a problem of direct current interest to society as a whole.

J. References

VI-1. Green, A. E. S.; and Wyatt, P. J: Atomic and Space Physics. Addison Wesley Publishing Co., Reading, Mass., 1965.

VI-2. McCormac, B. M.; and Omholt, A.: Atmospheric Emissions. Van Nostrand Reinhold, New York, 1969.

VI-3. Chamberlain, J. W.: Physics of the Aurora and Airglow. Academic Press, New York, 1961.

VI-4. Ratcliffe, J. A.: Physics of the Upper Atmosphere. Academic Press, New York, 1960.

VI-5. Whitten, R. C.; and Poppoff, I. G.: Physics of the Lower Ionosphere. Prentice Hall, Englewood Cliffs, N. J., 1965.

VI-6. Green, A. E. S.; and Barth, C. A.: Calculations of ultraviolet nitrogen emissions from the aurora. J. Geophys. Res., vol. 70, 1965, p. 1083.

VI-7. Stolarski, R. S.; and Green, A. E. S.: Calculations of auroral intensities from electron impact. J Geophys. Res. vol 72, 1967, p. 3967.

VI-8. Green, A. E. S.; and Barth, C. A.: Calculations of the photoelectron excitation of the dayglow. J. Geophys. Res., vol 72, 1967, p. 3975.

VI-9. Peterson, L. R.; and Green, A. E. S.: The relation between ionization yields, cross sections and loss functions. J. Phys. B (Proc. Phys. Soc.), vol. 1, 1968, p. 1131.

VI-10. Dalgarno, A.; McElroy, M. B.; and Stewart, A. I.: Electron impact excitation of the dayglow. J. Atmos. Sci., vol. 26, 1969, p. 753.

VI-11. Green, A. E. S.; Peterson, L. R.; and Prasad, S. S.: Semi-empirical cross sections and the airglow and aurora. In Atmospheric Emissions, ed. by B. M. McCormac and A. Omholt, Van Nostrand Reinhold, New York, 1969, pp. 523-532.

VI-12. Evans, W. F. J.; and Llewellyn, E. J.: Molecular oxygen emissions in the airglow. Ann. Geophys., vol. 26, 1970, p. 167.

VI-13. Noxon, J. F.: Auroral emission from $O_2(^1\Delta_g)$. J. Geophys. Res., vol. 75, 1970, p. 1879.

VI-14. Megill, L. R.; Despain, A. M.; Baker, D. J.; and Baker, K. D.: Oxygen atmospheric and infrared atmospheric bands in the aurora. J. Geophys. Res., vol. 75, 1970, p. 4775.

VI-15. Stolarski, R. S.; Dulock, V. A., Jr.; Watson, C. E.; and Green, A. E. S.: Electron impact cross sections for atmospheric species, 2. molecular nitrogen. J. Geophys. Res., vol. 72, 1967, p. 3953; Watson, C. E.; Dulock, V. A., Jr.; Stolarski, R. S.; and Green, A. E. S.: Electron impact cross sections for atmospheric species, 3. molecular oxygen. J. Geophys. Res., vol. 72, 1967, p. 3961.

VI-16. Peterson, L. R.; Prasad, S. S.; and Green, A. E. S.: Semi-empirical electron impact cross sections for atmospheric gases. Can. J. Chem., vol. 47, 1969, p. 1774.

VI-17. Lassettre, E. N.: Inelastic scattering of high energy electrons by atmospheric gases. Can. J. Chem., vol. 47, 1969, p. 1733.

VI-18. Geiger, J.; and Schroder, B.: Intensity perturbations due to the configuration interaction observed in the electron energy-loss spectrum of N_2. J. Chem. Phys., vol. 50, 1969, p. 7.

VI-19. Williams, A. J., III; and Doering, J. P.: Low-energy electron-impact study of the 12-14 eV transitions in nitrogen. J. Chem. Phys., vol. 51, 1969, p. 2859.

VI-20. Brinkman, R. T.; and Trajmar, S.: Electron impact excitation of N_2. Ann. Geophys., vol. 26, 1970, p. 201.

VI-21. Henry, R. J. W.; Burke, P. G.; and Sinfailam, A. L.: Scattering of electrons by C, N, O, N^+, O^+, and O^{++}. Phys. Rev., vol. 178, 1969, p. 218.

VI-22. Schultz, G. J.: Vibrational excitation of N_2, CO and H_2 by electron impact. Phys. Rev., vol. 135, 1964, p. A988.

VI-23. Butler, S. T.; and Buckingham, M. J.: Energy loss of a fast ion in a plasma. Phys. Rev., vol. 126, 1962, p. 1.

VI-24. Hake, R. D.; and Phelps, A. V.: Momentum-transfer and inelastic collision cross sections for electrons in O_2, CO and CO_2. Phys. Rev., vol. 158, 1967, p. 70.

VI-25. Schultz, G. J.; and Dowell, J. T.: Excitation of vibrational and electronic levels in O_2 by electron impact. Phys. Rev., vol. 128, 1962, p. 174.

VI-26. Borst, W. L.; and Zipf, E. C.: Cross section for electron-impact excitation of the (o, o) first negative band of N_2^+ from threshold to 3 keV. Phys. Rev., vol. A1, 1970, p. 834.

VI-27. Kieffer, L. J.; and Dunn, G. H.: Electron impact ionization cross-section data for atoms, atomic ions, and diatomic molecules: I. experimental data. Rev. Mod. Phys., vol. 38, 1966, p. 1.

VI-28. Ajello, J. M.: Emission cross sections of N_2 in the vacuum ultraviolet by electron impact. J. Chem. Phys., vol. 53, 1970, p. 1156.

VI-29. Herzberg, G.: Molecular Spectra and Molecular Structure I. Spectra of Diatomic Molecules. D. Van Nostrand Co., Princeton, N. J., 1950.

VI-30. Dunn, G. H.: Anisotropies in angular distributions of molecular dissociation products. Phys. Rev. Letters, vol. 8, 1962, p. 62.

VI-31. van Brunt, R. J.; and Zare, R. N.: Polarization of atomic fluorescence excited by molecular dissociation. J. Chem. Phys., vol. 48, 1968, p. 4304.

VI-32. Leventhal, M.; Robiscoe, R. T.; and Lea, K. R.: Velocity distribution of metastable H atoms produced by dissociative excitation of H_2. Phys. Rev., vol. 158, 1967, p. 49.

VI-33. Sroka, W.: Lichtemission im vakuumultraviolett durch electronenstossanregung in Gasen, teil A: untersuchungen in sauerstoff. Z. Naturforsch., vol. 23, 1968, p. 2004; Sroka, W.: Lichemission in vakuumultraviolett durch electronenstossanregung in gasen, teil B: untersuchungen in stickstoff. Z. Naturforsch., vol. 24, 1969, p. 398.

VI-34. Mumma, M. J.: Dissociative excitation of atmospheric gases. Ph.D. Dissertation, University of Pittsburgh, 1970.

VI-35. Ajello, J. M.: Emission cross sections of N_2 in the vacuum ultraviolet by electron impact. Ph.D. Thesis, University of Colorado, 1969.

VI-36. Aarts, J. F. M.; de Heer, F. J.; and Vriens, L.: Excitation of N_2 by 50 - 5000 eV electrons. In Proc. VIth ICPEAC, MIT Press, Cambridge, Mass., 1969, p. 423.

VI-37. Dunn, G. H.; and Kieffer, L. J.: Dissociative ionization of H_2: a study of angular distributions and energy distributions of resultant fast protons. Phys. Rev., vol. 132, 1963, p. 2109.

VI-38. Walker, J. C. G.; and Rees, M. H.: Ionospheric electron densities and temperatures in aurora. Planet. Space Sci., vol. 16, 1968, p. 459.

VI-39. Dalgarno, A.: Inelastic collisions at low energies. Can. J. Chem., vol. 47, 1969, p. 1723.

VI-40. COSPAR International Reference Atmosphere, 1965. Compiled by COSPAR working group IV, North Holland Publishing Company, Amsterdam, 1965.

VI-41. Evans, W. F. J.; Hunten, D. M.; Llewellyn, E. J.; and Jones, A. Vallance: Altitude profile of the infrared atmospheric system of oxygen in the dayglow. J. Geophys. Res., vol. 73, 1968, p. 2885.

VI-42. Wallace, L.; and Hunten, D. M.: The dayglow of the oxygen A band. J. Geophys. Res., vol. 73, 1968, p. 4813.

VI-43. Schiff, H. I.; Haslett, J. C.; and Megill, L. R.: Accidental rocket observation of $O_2(^1\Delta_g)$ in an aurora. J. Geophys. Res., vol. 75, 1970, p. 4363.

VI-44. Isaksen, I.; Llewellyn, E. J.; and Jones, A. Vallance: Photo-chemical effects in the atmosphere resulting from auroral bombardment. Institute of Space and Atmospheric Studies, University of Saskatchewan, Report No. 2, 1968.

VI-45. Lin, C. L.; and Kaufman, F.: Deactivation and reaction of $N(^2D)$. Presented at the DASA symposium on the Physics and Chemistry of the Upper Atmosphere, Stanford Research Institute, Menlo Park, California, June 23-24, 1969.

VI-46. Black, G.; Slanger, T. G.; St. John, G. A.; and Young, R. A.: Vacuum-ultraviolet photolysis of N_2O. IV deactivation of $N(^2D)$. J. Chem. Phys., vol. 51, 1969, p. 116.

VI-47. Noxon, J. F.: Optical emission from $O(^1D)$ and $O_2(b^1\Sigma_g)$ in ultraviolet photolysis of O_2 and CO_2. J. Chem. Phys., vol. 52, 1970, p. 1852.

VI-48. Wayne, R. P.: Laboratory studies on the excitation and deactivation of singlet molecular oxygen. Annals N. Y. Acad. Sci., vol. 171, 1970, p. 199; Clark, I. D.: Energy transfer from $O(^1D)$ to O_2. Chem. Phys. Letters, vol. 5, 1970, p. 317.

VI-49. Walker, J. C. G.: Electric field excitation of $O_2(^1\Delta_g)$ in auroras. Planet. Space Sci., vol. 18, 1970, p. 1043.

VI-50 Berthelier, J. J.; and Sturges, D. J.: Simultaneous measurements of electron density and temperature in the northern auroral zone. Planet. Space Sci., vol. 15, 1967, p. 1049.

VI-51. Omholt, A.: Studies on the excitation of aurora Borealis II. The forbidden lines of oxygen. Geofys. Publik. (Norway), vol. 21, no. 1, 1959.

VI-52. Rees, M. H.; and Walker, J. C. G.: Ion and electron heating by auroral electric fields. Ann. Geophys., vol. 24, 1968, p. 193.

VI-53. Roble, R. G.; and Dickinson, R. E.: Atmospheric response to heating within a stable auroral red arc. Planet. Space Sci., vol. 18, 1970, p. 1489.

VI-54. Katz, R.: Track theory and radiation quality. Proceedings of the Symposium on Biophysical Aspects of Radiation Quality sponsored by the International Atomic Energy Agency, Lucas Heights, Australia, March 8-12, 1971.

VI-55. Turner, J. E.: Meaning and assessment of radiation quality for radiation protection. Proceedings of the Symposium on Biophysical Aspects of Radiation Quality sponsored by the International Atomic Energy Agency, Lucas Heights, Australia, March 8-12, 1971.

VI-56. Green, A. E. S.; Olivero, J. J.; and Stagat, R. W.: Microdosimetry of Low Energy Electrons. Proceedings of the Symposium of Biophysical Aspects of Radiation Quality sponsored by the International Atomic Energy Agency, Lucas Heights, Australia, March 8-12, 1971.

VI-57. Rossi, H. H.: Dual radiation injury and its consequences. Proceedings of the Symposium on Biophysical Aspects of Radiation Quality sponsored by the International Atomic Energy Agency, Lucas Heights, Australia, March 8-12, 1971.

VI-58. Green, A. E. S.; Sellin, D. L.; and Zachor, A. S.: Phys. Rev., vol. 184, 1969, p. 1.

VI-59. Ganas, P. S.; Dutta, S. K.; and Green, A. E. S.: Phys. Rev. A, vol. 2, 1970, p. 1.

VI-60. Purcell, J. E.; Berg, R. A.; and Green, A. E. S.: Phys. Rev. A, vol. 2, 1970, p. 1.

VI-61. International Conference on Singlet Molecular Oxygen and Its Role in Environmental Sciences. Annals of the New York Academy of Sciences, vol. 171, no. 1, 1970.

VI-62. Schiff, H. I.: Role of singlet oxygen in upper atmosphere chemistry. International Conference on Singlet Molecular Oxygen and Its Role in Environmental Sciences, vol. 171, no. 1, 1970, 1. 188.

VI-63. Myers, G. H.; and O'Brien, R. J. Jr.: Quenching of $O_2(b^1\Sigma_g^+)$. International Conference on Singlet Molecular Oxygen and Its Role in Environmental Sciences, vol. 171, no. 1, 1970, p. 224.

K. Bibliography

Ganas, P. S.; and Green, A. E. S.: Electron impact excitation of the rare gases. Phys. Rev., To be published.

Green, A. E. S.; and Strickland, D. J.: Atomic processes in the polar cap. To be published in special issue on the polar cap, J. Franklin Inst., 1971.

Prasad, S. S.; and Green, A. E. S.: Ultraviolet emissions from atomic nitrogen in the aurora. To be published in J. Geophys. Res.

Sawada, T.; Purcell, J. E.; and Green, A. E. S.: Distorted wave calculations of electron impact excitation of the rare gases. Phys. Rev., To be published.

CHAPTER VII. THE COMPOSITION OF THE UPPER ATMOSPHERE

By

Alfred O. Nier
School of Physics and Astronomy
University of Minnesota, Minneapolis, Minnesota

A. Abstract

Miniature mass spectrometers have been developed and have been carried on sounding rockets to determine the composition of the upper atmosphere. Techniques have been developed that accurately correct for the velocity and spin of the moving vehicle. Above 120 km N_2, O_2, and Ar appear to be in diffusive equilibrium. Most He concentration measurements show a more rapid decline with altitude than predicted by diffusive equilibrium. Because of the highly reactive nature of atomic oxygen, measurements of this species by mass spectrometry are low by an unknown factor. New experiments promise to supply information on this question.

B. Introduction

Mass spectrometers have been used in the research laboratory and in industry for analyzing complex mixtures of gases since approximately 1940. The instruments are large and heavy, and consume considerable power. After World War II when sounding rockets became available and exploration of the upper atmosphere was begun, miniaturized mass spectrometers carried on rockets began to play an important part in determining the composition of the upper atmosphere. Neutral analyses were made for the first time in 1953 [VII-1] and ion analyses in 1954 [VII-2].

Many improvements have been made in the years since, and various types of instruments have been used, on satellites as well as on rockets. These include radio-frequency Bennett instruments [VII-3], quadrupoles [VII-4-VII-6], monopoles [VII-7], omegatrons [VII-8], and several types of magnetic deflection instruments [VII-9-VII-11]. This paper will discuss some of the results obtained with magnetic deflection instruments carried on sounding rockets launched by the University of Minnesota group at the White Sands, New Mexico, missile range and at Fort Churchill, Manitoba, Canada.

C. Instrumentation

In all of the studies to be discussed here the instruments were carried just inside the cylindrical skin of the rocket. Figure VII-1 is a sectional view of a typical instrument. Ions are produced by an electron beam perpendicular to the figure. The ions formed are drawn out and accelerated between several focusing plates before passing into the space

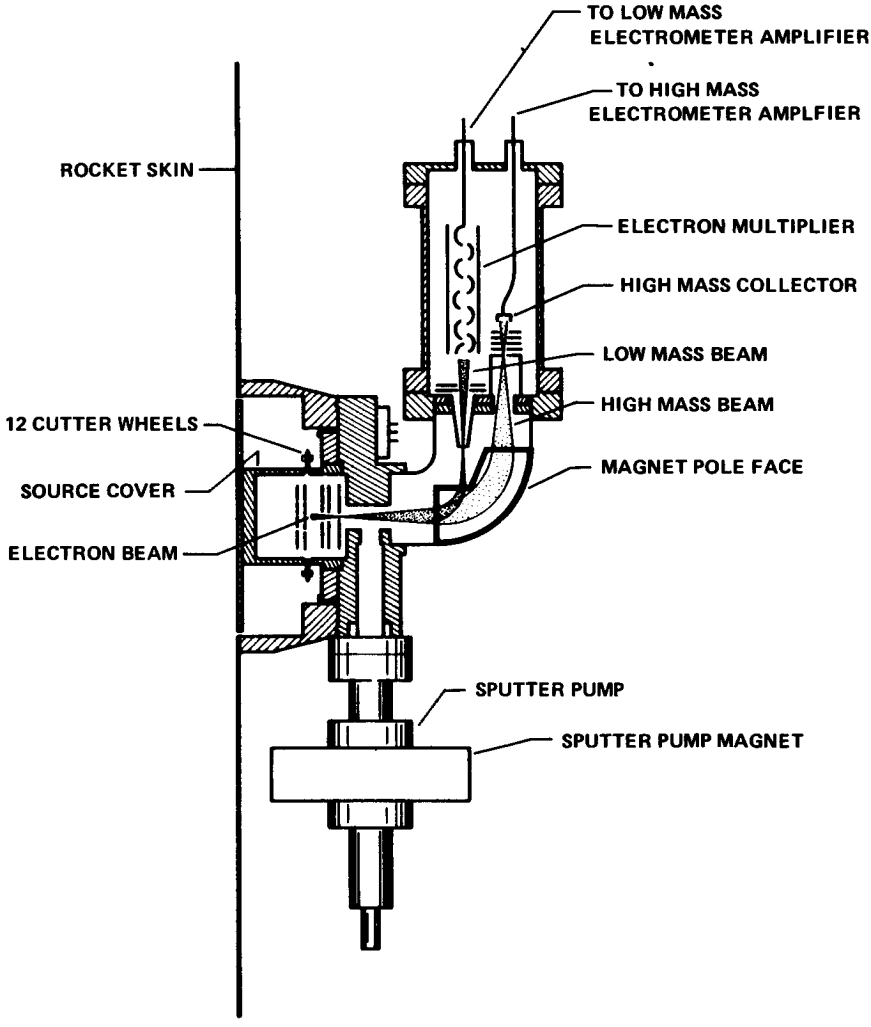


Figure VII-1. Schematic drawing of a typical magnetic deflection mass spectrometer carried on Aerobee rockets for making measurements in the 120- to 200-km altitude range.

between the poles of a magnet. The magnetic field deflects the ions and also focuses a diverging beam as illustrated. A mass spectrum is swept by varying the ion accelerating voltage, the period of sweep being approximately 2 sec. As a practical matter two collectors are employed; one collects ions in the approximate mass range of 12 to 48 amu and the other in the range of 3 to 12 amu. This has the advantage that one collector does not cover a very wide range. Moreover, an electron multiplier pre-amplifier, providing a gain of about 20 000, can be employed in conjunction with the light mass collector to measure the low abundance ions, He^+, N^{2+}, and O^{2+}.

The mass spectrometer is calibrated in the laboratory and sealed and pumped with a small ion sputter pump. The small residual gas remaining provides ions for test during rocket integration and before flight. An aluminum cap, covering the ion source, is cut off by a pyrotechnic-operated mechanism [VII-12] when the rocket reaches an altitude of approximately 100 km. The ion source is then exposed, and because of its "open" nature, ambient gas can enter the ionizing region directly with a minimum of collisions.

The open source, while useful for measuring reactive species such as atomic oxygen, introduces some problems. A gas molecule striking a wall of the source will rebound with a speed that depends upon the accommodation coefficient of the surface. The gas particle density in the path of the ionizing electron beam is determined by the sum of those particles that come directly from the atmosphere without making wall collisions and those that have made wall collisions. The number density in the reflected beam depends upon the speed of the particles and hence upon the accommodation coefficient. Because of this uncertainty, as well as other factors, it was felt desirable to simultaneously fly "closed" source instruments as shown in the cross-sectional view of the rocket shown in Figure VII-2. The no. 2 instrument is identical to the open source no. 1 instrument. The no. 3 spectrometer scans only over the mass 28 peak, at a rate of 7 times per second, and hence for all practical purposes continuously monitors the molecular nitrogen. Both the no. 2 and no. 3 instruments are attached to a common cavity which in turn connects to the ambient atmosphere through a knife-edged hole. As a result there is a calculable relation between the particle density of a nonreactive species in the cavity and in the ambient atmosphere.

D. Data

Figure VII-3 is a section of a typical telemetry record showing the outputs of the three mass spectrometers as well as the magnetometers and solar sensors for NASA Aerobee rocket flight 4.180 UA launched at White Sands Missile Range at 1409 MST on December 2, 1966 [VII-13].

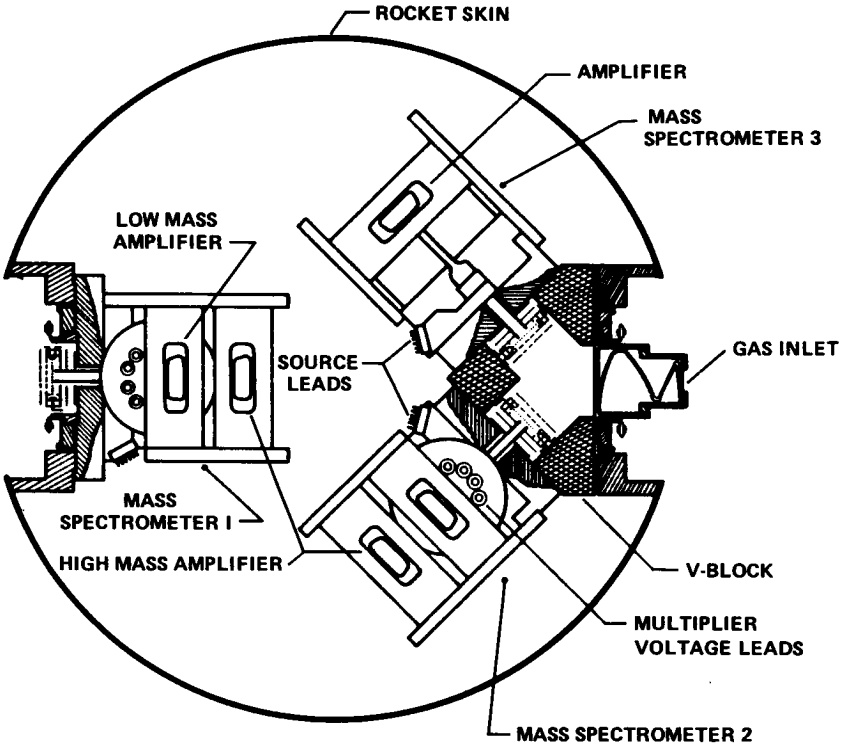


Figure VII-2. Cross section of a typical payload showing arrangement of three mass spectrometers.

Figure VII-4 is a plot of the intensity of the mass 28 peaks as observed with the no. 3 instrument as the rocket rises from 100 to 200 km and drops down again to 100 km. The modulation of the data is caused by the spin of the rocket as the gas inlet alternately looks into the "ram" and "wake" directions. The reduction of the raw data to ambient particle densities has been discussed elsewhere [VII-14] and will not be repeated here. It suffices to say that an equation for the curve can be expressed in terms of ambient N_2 density, temperature, rocket spin, aspect, and velocity components. A least-squares analysis provides the "best" values for the various parameters and gives the variation of N_2 density and atmosphere temperature as a function of altitude.

Figure VII-5 shows the agreement for two arbitrarily chosen cycles between the raw data and the results of the least squares fit.

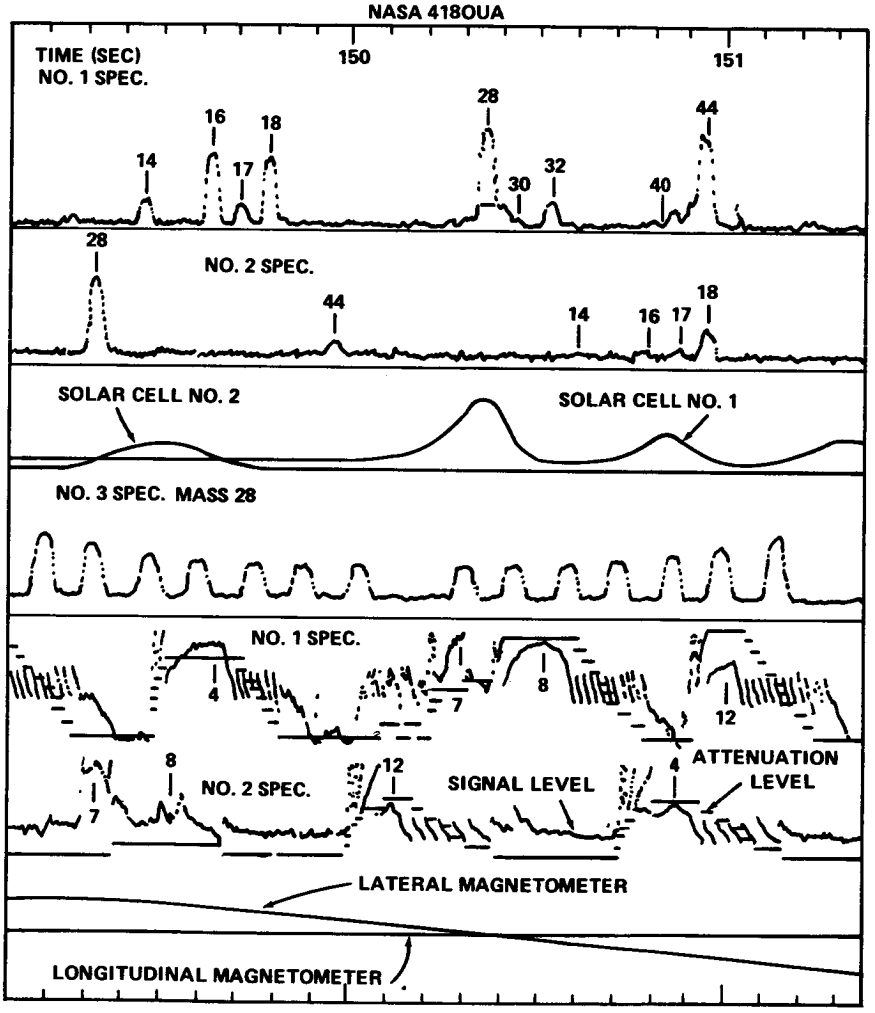


Figure VII-3. Section of telemetry record showing outputs of mass spectrometers, magnetometers, and aspect sensors (each of the electrometers had an automatic range changer that provided a reduction in sensitivity up to a factor of 1024 by factors of 2; the signal showing the range change step is printed just below the spectrum; in the case of the high mass signals from the no. 1 and no. 2 instruments (top of figure), only the 28 peak in the no. 1 spectrometer required an attenuation to remain on scale; in the case of the low mass amplifiers (near bottom of figure), many changes in range are evident).

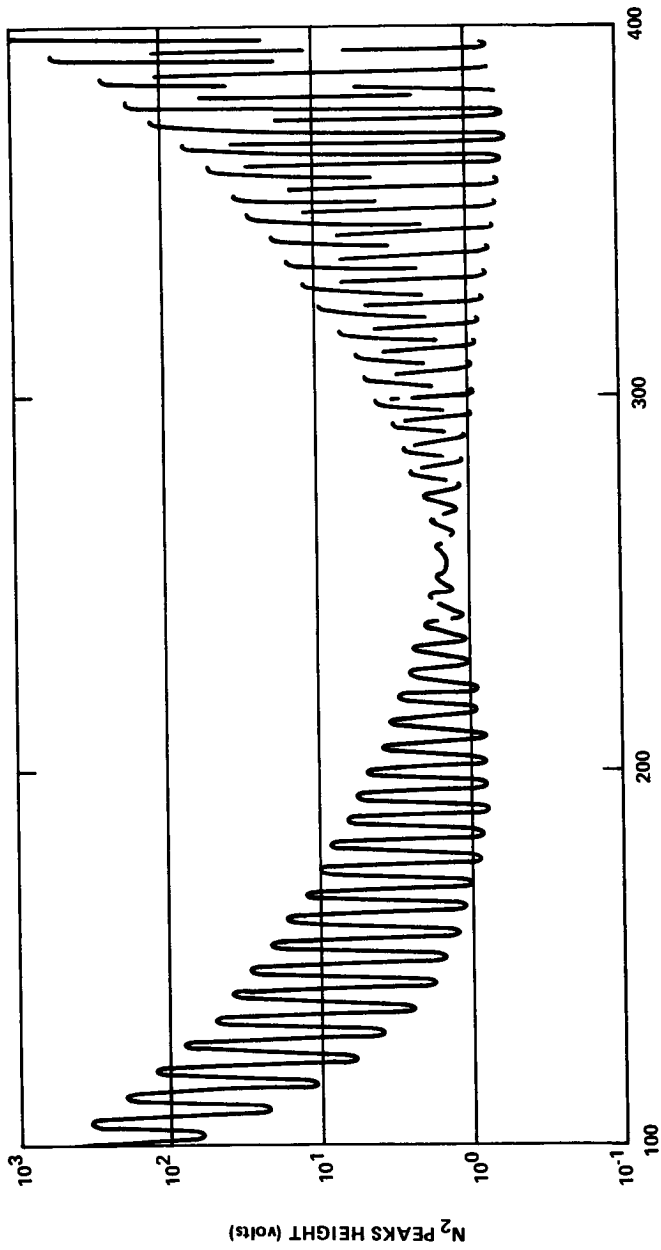


Figure VII-4. Mass 28 signal in no. 3 instrument of flight 4.180 UA showing the effect of spin of the rocket on the measurements (the breaks in the curve on the downleg part of flight are caused by a malfunction of the radio transmitter on this particular flight; this had no appreciable effect on the quality of the data since the curve already had so many points (over 2000) that data had to be discarded to reduce the computing load).

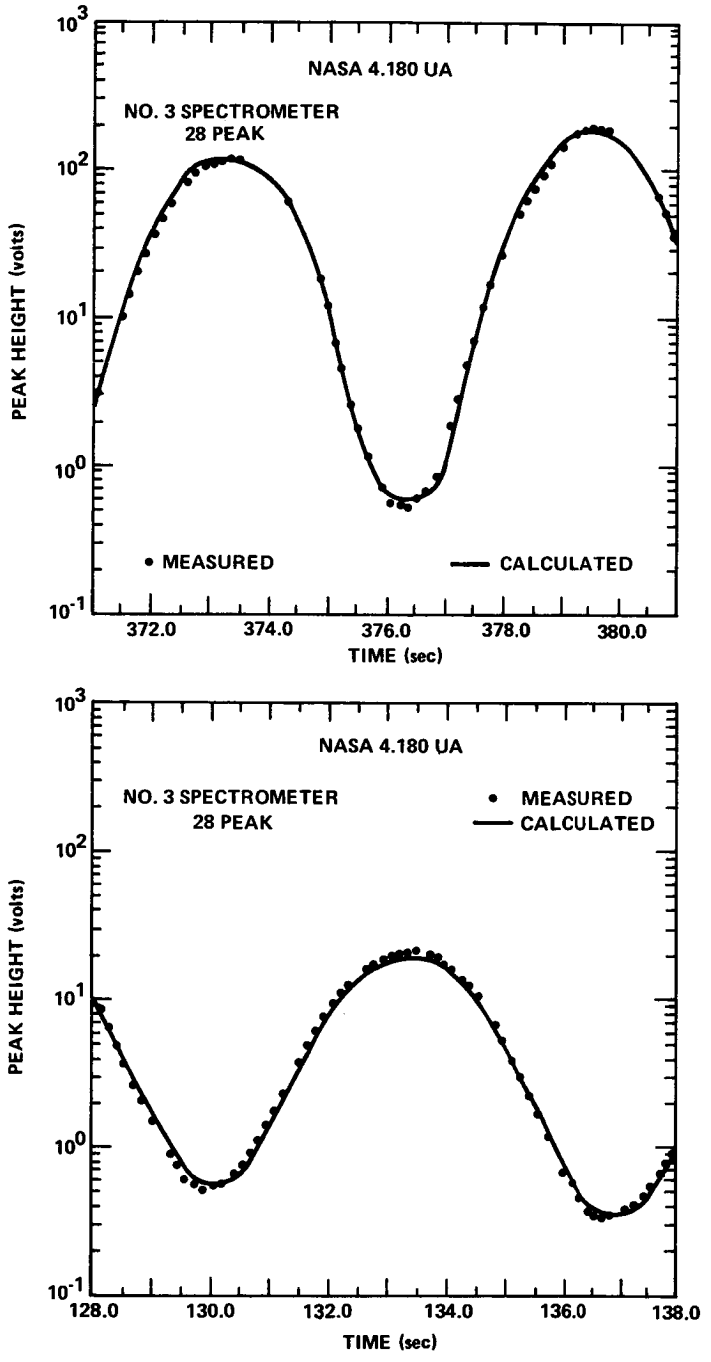


Figure VII-5. Detailed section of two cycles of curve showing agreement between raw data and theoretical curve based on a least-squares fitting process.

E. Analysis

The parameters found from the no. 3 instrument N_2 analyses are used to compute the number densities of the other major constituents, O_2, O, Ar, and He. Except for atomic oxygen where one expects a difference, the no. 1 and no. 2 instruments agree fairly well, showing that the "open" instrument is actually fairly "closed" and gives reasonably quantitative number densities.

In Figure VII-6 are plotted the computed number densities as a function of altitude. Also shown are the variations with altitude expected from the vertical temperature profile on the assumption that diffusive equilibrium prevails. For convenience the curves are normalized at 150 km. To demonstrate the insensitivity of the results to the exact absolute temperature assumed at the top of the flight, points are shown based on the assumption that the temperature at 220 km is 700° K and 900° K, respectively. It is seen that, except for helium, the constituents are indeed in diffusive equilibrium. This result has been confirmed in numerous other similar flights [VII-15-VII-17].

Figure VII-7 gives a summary of the results obtained for N_2, O_2, and Ar for seven different flights at White Sands and two at Fort Churchill. Except for flight 4.212, where the curves all lie low, there is remarkably good agreement between the results of the several flights. The low values at high altitudes for flight NC 3.115F are almost certainly a result of the lower atmosphere temperature, and hence shorter scale heights, at the time.

While the deviation of the 4.212 results could be attributed to an atmosphere variation, it seems most reasonable to assume it was caused by some instrument malfunction that did not affect relative readings. A change in the high mass electrometer input resistor after calibration could lead to such a result. While some of the variations between other flights can undoubtedly be attributed to calibration or instrument errors, the redundancy [VII-13, VII-16] provided by these instruments making simultaneous measurements suggests that at least part of the variations observed are true atmospheric fluctuations.

F. Helium

As can be seen from Figure VII-6, helium poses a special problem. Whereas, in this flight, all of the other components appear to be in diffusive equilibrium above 120 km, this is decidedly not true for helium. This constituent clearly declines with altitude faster than predicted by diffusive equilibrium. Moreover, as evident in Figure VII-8, the absolute abundance varies by a factor of more than 10, with the winter White Sands flights

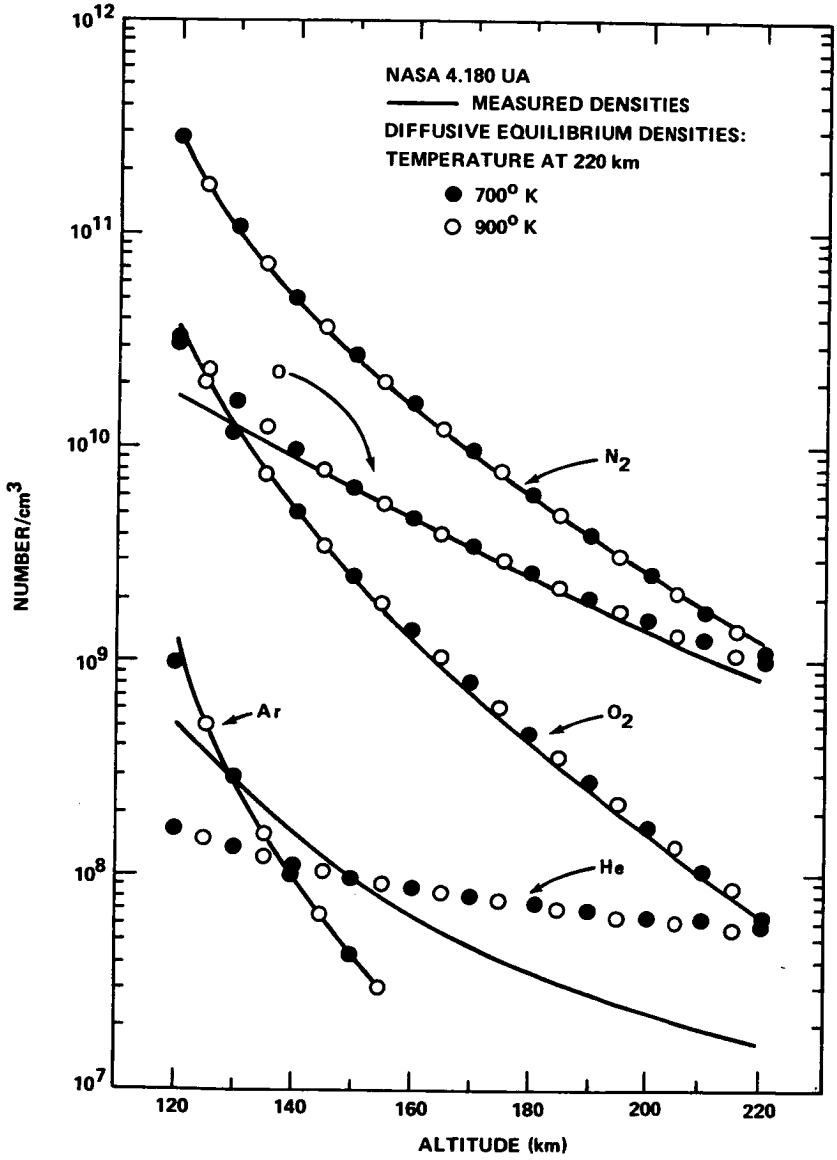


Figure VII-6. Number densities for N_2, O_2, O, Ar, and He found for flight 4.180 UA (the points correspond to curves computed assuming that diffusive equilibrium prevails).

giving a much higher abundance than the corresponding spring and summer flights. Winter and spring Fort Churchill values fall between these values.

In an analysis of seven White Sands flights, Kasprzak [VII-18] showed that a possible explanation for the deviation from diffusive equilibrium in

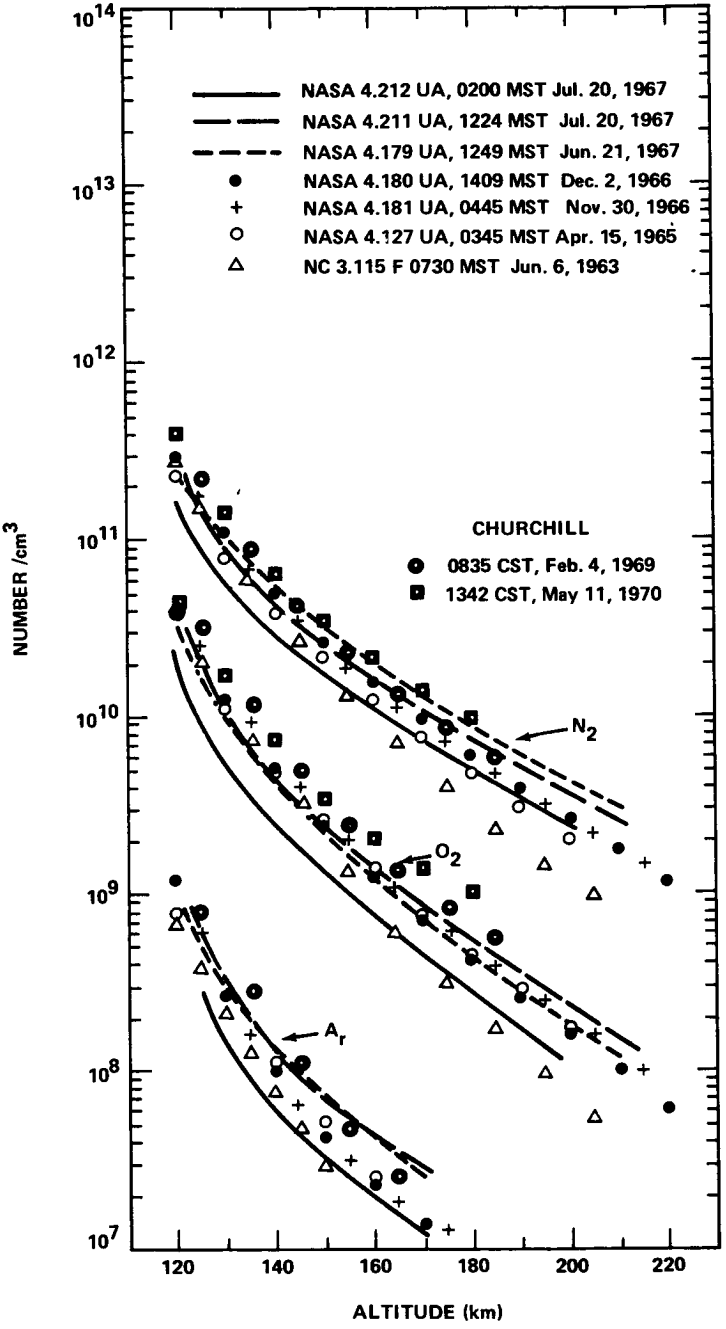


Figure VII-7. Number densities as a function of altitude for N_2, O_2, and Ar for several flights.

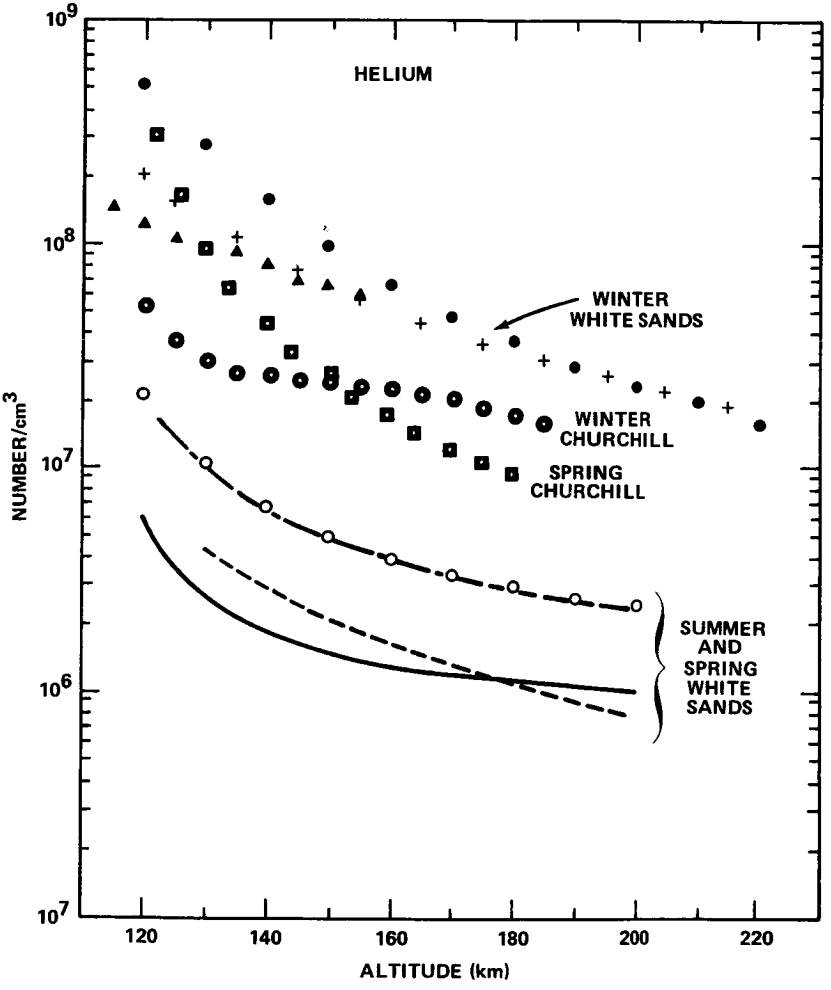


Figure VII-8. Number density as a function of altitude for He for the same flights shown in Figure VII-7 (symbols in the two figures are consistent; also shown for comparison (solid triangles) are the December 12, 1966, Churchill results of Hartmann et al. [VII-19].

the lower thermosphere could be found if one postulated an appropriate upward flux of helium. Hodges [VII-20] has shown that an alternate explanation could be found for behavior of a minor constituent, such as helium, if gravity waves were present.

A spring flight at Fort Churchill[1], NASA 4.322 UA, showed the same steeper-than-diffusive equilibrium variation exhibited by all

1. Unpublished work of A. O. Nier.

White Sands flights; whereas, a winter flight [VII-21], NASA 4.272 UA, agreed very closely with diffusive equilibrium, which agreed with an earlier result of Hartman et al. [VII-19].

That the concentration variations are somehow related to dynamic effects such as the winter polar helium bulge [VII-22] seems very likely. The explanation for the deviations from diffusive equilibrium in the lower thermosphere is not clear at present, but the deviations must surely represent some dynamic effect in the atmosphere.

G. Atomic Oxygen

There is little reason to doubt the absolute concentration values found for nonreactive gases such as N_2, O_2, Ar, and He, but the same cannot be said for atomic oxygen. In spite of the open nature of the ion source shown in Figure VII-1, an appreciable portion of the particle number density occurring in the ionizing electron beam is a result of wall collisions. If any sizable part of the atomic oxygen is lost through such collisions, the atomic oxygen readings will clearly be low compared to the nonreactive gases. As seen in Figure VII-6 and in corresponding graphs for other flights, atomic oxygen measurements are consistent with this constituent being in diffusive equilibrium. Also, when curves similar to Figure VII-4 are constructed for atomic oxygen, the spin modulation data and the comparison between upleg and downleg data suggest that if atomic oxygen is lost as a result of surface collisions, such loss is independent of time (during the flight) and of intensity of the signal. In other words, in rocket flights such as those discussed here, atomic oxygen behaves as does any nonreactive gas except that all readings may be low by a constant factor in a particular flight instrument.

In an effort to evaluate this factor, use has been made of the fact that in the rocket flights discussed, there have been both open and closed ion source mass spectrometers. The open source instruments always show an appreciable amount of atomic oxygen, and the closed source instruments show only a small amount. Kasprazak et al. [VII-13] made use of this observation in their attempt to evaluate the atomic oxygen loss. Briefly, the method was as follows: after correcting the mass 16 peak in the spectra for the fragment O^+, due to the dissociation of O_2 by the ionizing electron beam, the remainder is attributed to the ionization of atomic oxygen. Figure VII-9 shows the results for a typical flight, NASA 4.180 UA. Here we see that the closed source instrument shows about 5 percent as much O as the open source instrument. From a knowledge of the relative surface areas in the two sources and a Monte Carlo calculation of the number of collisions an average atomic oxygen atom makes

before being ionized, one can estimate the probability of an atomic oxygen atom being lost in a collision. As seen from Figure VII-9 this probability turned out to be 12 percent. Using this number, one can then correct the atomic oxygen as read by the open (or closed) source instrument. The

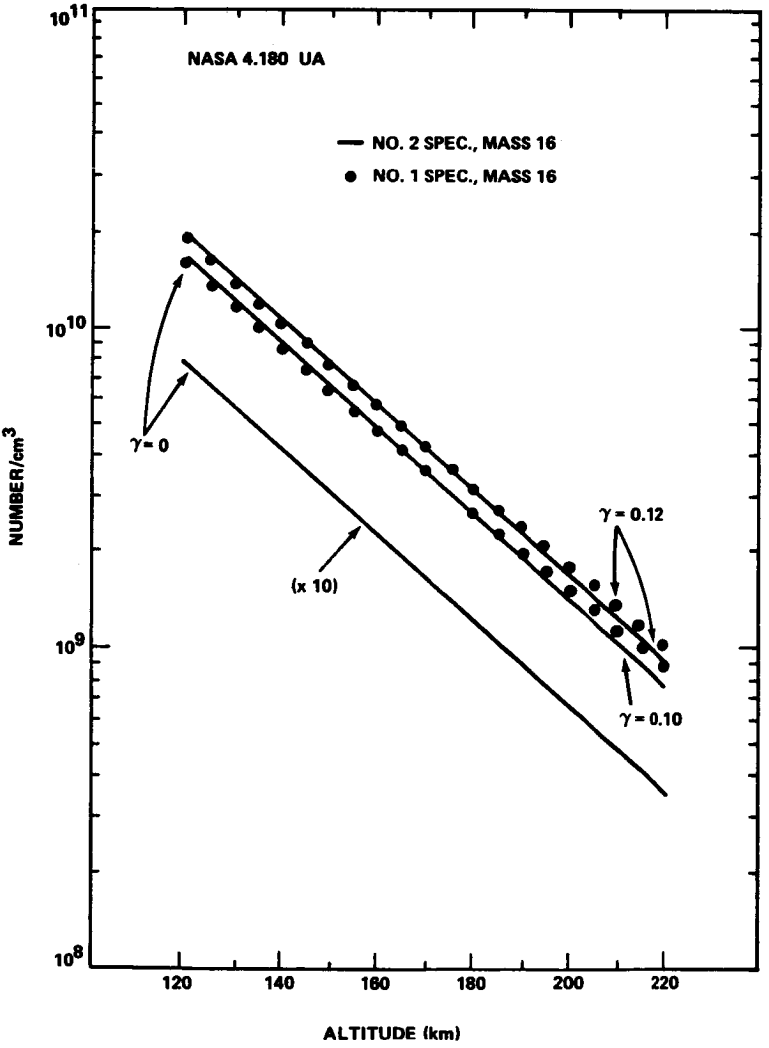


Figure VII-9. Comparison of atomic oxygen measurements as found with open (no. 1) and closed (no. 2) instruments in flight 4.180 UA (curves marked $\gamma = 0$ are for data uncorrected for O loss; $\gamma = 0.10$ curve shows value of loss coefficient necessary to make no. 2 data agree with no. 1 instrument; curves marked $\gamma = 0.12$ are obtained if it is assumed that losses occur in both instruments).

conclusion is that the open source atomic oxygen reading is low by about 20 percent. Similar analyses applied to other flights led to approximately the same conclusions.

Figure VII-10 gives a summary for eight flights by the Minnesota group. The numbers are uncorrected for the effect discussed above and should be multiplied by an appropriate factor, 1.25, if one subscribes to the analyses given. It is seen that there are wide variations from flight to flight, even ignoring the 0200, July 20, 1969 flight which, as mentioned earlier, may have had an error in calibration. While one might be tempted

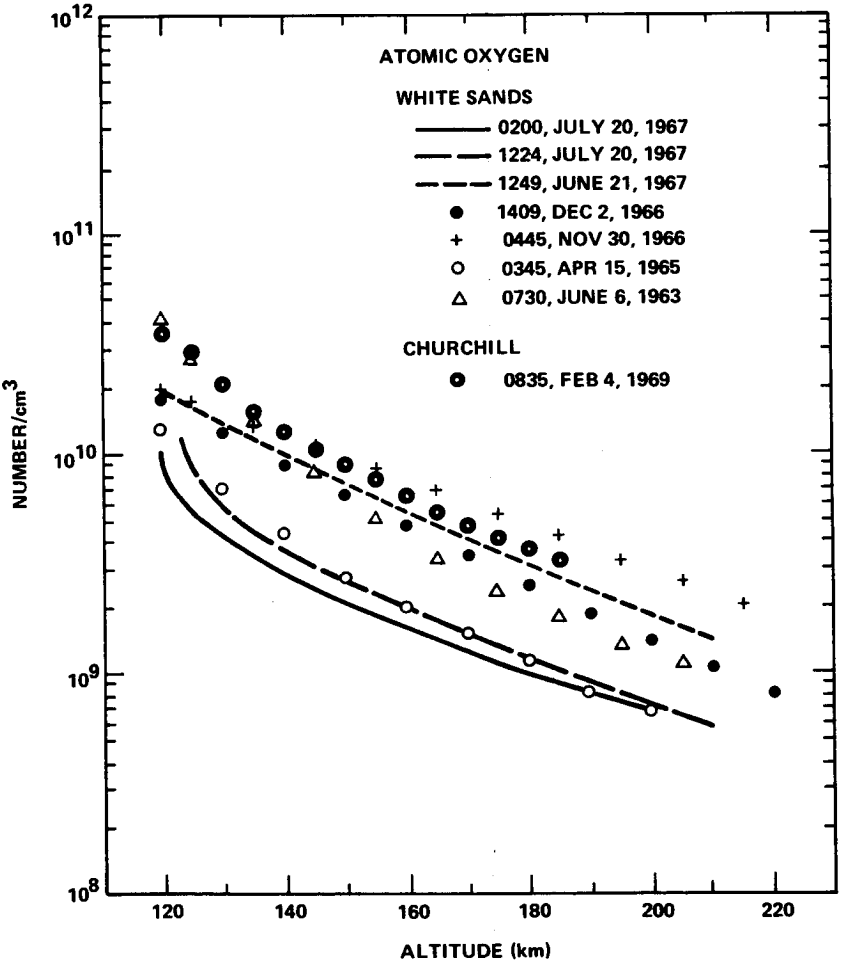


Figure VII-10. Comparison of atomic oxygen values obtained for the several flights shown in Figures VII-7 and VII-8.

to attribute the differences to variations in the atomic oxygen loss in the ion sources, it seems more plausible, without evidence to the contrary, to associate the variations with real geophysical effects, especially since other types of experiments suggest that the atomic oxygen concentration does vary in the thermosphere.

In an interesting discussion of the atomic oxygen problem, von Zahn [VII-23] made a comparison of all mass spectrometer data at 150 km with extreme ultraviolet extinction measurements and satellite drag measurements. He concluded that the various results could be brought into agreement if one slightly reduced the densities at 150 km deduced from satellite drag and assumed that all rocket-borne mass spectrometer atomic oxygen measurements were too low by an appreciable factor. If his deduction is correct, one would also have to reduce the atmospheric temperature used in current models, since a greatly increased atomic oxygen concentration at 150 km, when extrapolated to altitudes where atomic oxygen is the principal constituent, could result in atmospheric densities well above what are observed from satellite drag measurements.

In an ingenious experiment in which a mass spectrometer having a liquid helium cooled ion source was carried on a rocket, von Zahn found an O/O_2 ratio of 3.4 at 120 km. The experiment assumes that both O and O_2 are totally absorbed on contact with a cold surface, so one measures only those ambient particles that have not made collision. The result differs so markedly from other measurements that it should be checked by independent means.

Joint experiments are underway between the University of Minnesota group and Dr. J. B. French and his colleagues of the University of Toronto Institute of Aerospace Studies in which the performance of mass spectrometers will be studied in the Toronto molecular beam facility. The availability of calibrated beams, including atomic oxygen, should elucidate many questions, especially the oxygen loss problem.

In conclusion it may be said that mass spectrometry has advanced to the point where, with techniques already developed or underway, it should be possible to make relatively precise measurements of atmospheric composition and, hence, of density.

H. References

VII-1. Townsend, J. W., Jr.; Meadows, E. B.; and Pressley, E.: Rocket Exploration of the Upper Atmosphere. Edited by R. L. F. Boyd and M. J. Seaton, Interscience Publishers, New York.

VII-2. Johnson, C. Y.; and Meadows, E. B.: J. Geophys. Res. vol. 60, 1955, p. 193.

VII-3. Bennett, W. H.: J. Appl. Phys., vol. 21, 1950, p. 143.

VII-4. Schaefer, E. J.; and Nichols, M. H.: J. Geophys. Res. vol. 69, 1964, p. 4649.

VII-5. Mauersberger, K.; Müller, D.; Offerman, D.; and von Zahn, U.: J. Geophys. Res., vol. 73, 1968, p. 1071.

VII-6. Müller, D.; and Hartmann, G.: J. Geophys. Res., vol. 74, 1969, p. 1287.

VII-7. von Zahn, U.; and Gross, J.: J., Geophys. Res., vol. 74, 1969, p. 4055.

VII-8. Spencer, N. W.; Taeusch, D. R.; and Carignan, G. R.: Ann. de Geophys., vol. 22, 1966, p. 151.

VII-9. Spencer, N. W.; and Reber, C. A.: Space Research, vol. 3, 1963, p. 1151.

VII-10. Nier, A. O.; Hoffman, J. H.; Johnson, C. Y., and Holmes, J. C.: J. Geophys. Res., vol. 69, 1964, p. 979.

VII-11. Hoffman, J. H.: J. Geophys. Res., vol. 72, 1967, p. 1883.

VII-12. Thorness, R. B.; and Nier, A. O.: Rev. Sci. Instr., vol. 33, 1962, p. 1005.

VII-13. Kasprzak, W. T.; Krankowsky, D.; and Nier, A. O.: J. Geophys. Res., vol. 73, 1968, p. 6765.

VII-14. Hedin, A. E.; Avery, C. P.; and Tschetter, C. D.: J. Geophys. Res., vol. 69, 1964, p. 4637.

VII-15. Hedin, A. E.; and Nier, A. O.: J. Geophys. Res., vol. 70, 1965, p. 1273.

VII-16. Hedin, A. E.; and Nier, A. O.: J. Geophys. Res., vol. 71, 1966, p. 4121.

VII-17. Krankowsky, D.; Kasprzak, W. T.; and Nier, A. O.: J. Geophys. Res., vol. 73, 1968, p. 7291.

VII-18. Kasprzak, W. T.: J. Geophys. Res., vol. 74, 1969, p. 894.

VII-19. Hartmann, G.; Mauersberger, K.; and Müller, D.: Space Research, vol. 8, 1968, p. 940.

VII-20. Hodges, R. R., Jr.: J. Geophys. Res., vol. 75, 1970, p. 4842.

VII-21. Hickman, D. R.; and Nier, A. O.: Trans. AGU, vol. 51, 1970, p. 378.

VII-22. Keating, G. M.; and Prior, E. J.: Space Research, vol. 8, 1968, p. 982.

VII-23. von Zahn, U.: J. Geophys. Res., vol. 75, 1970, p. 5517.

I. Bibliography

von Zahn, U.: J. Geophys. Res., In Press.

CHAPTER VIII. CHEMISTRY OF THE UPPER ATMOSPHERE AND THE ROLE OF TRANSPORT PROCESSES

By

Richard A. Craig
Florida State University, Tallahassee, Florida

A. Abstract

The important chemical reaction affecting the neutral composition of the upper stratosphere, mesosphere, and lower thermosphere is discussed. The role of atmospheric transport processes is emphasized and the problem of taking these into account is elucidated.

B. Introduction

Natural atmospheric gases in the lower atmosphere are mostly chemically inactive. Molecular oxygen, molecular nitrogen, water vapor, carbon dioxide, and argon simply coexist in a mixture at atmospheric temperatures. However, in the upper atmosphere in the presence of ultraviolet solar radiation, the dissociation and ionization of the first four of these lead to an extremely complex chemistry. This paper attempts to describe only a few of the first principles of this subject; to go into great detail would require a book.

As background, some general statements can be made:

1. For present purposes, only exothermic (or very slightly endothermic) reactions need to be considered. Interactions with high energy particles from outside the atmosphere (e.g., cosmic rays and auroral particles) are special topics not included here, because they do not significantly affect the worldwide neutral composition.

2. Two-body combinations involving emission of a photon are usually unimportant. Three-body reactions are often important below 100 km, but not at higher levels with lower densities.

3. The presence of free electrons and positive ions, although not in itself a topic of this paper, must be taken into account because of its effect on the neutral composition.

4. Metastable states sometimes play an important role, and often radiate because of the low density and absence of walls.

5. It has been found that chemical reactions in a motionless atmosphere cannot explain some observed aspects of composition. There are some problems where atmospheric transport processes are surely important, although poorly understood. We shall return to these near the end of this paper.

C. Dissociation Mechanisms

Since dissociation of the major gases is a necessary condition for the other reactions, it is appropriate to discuss how and at what altitudes this takes place. The first thing to consider is photodissociation. Table VIII-1 gives wavelength intervals in which significant photodissociation takes place. Information on photoionization is also included because of the importance of charged particles which will become evident later.

There is no significant photodissociation of N_2 in the upper atmosphere. On the other hand, photodissociation of O_2 plays a very important role. Strong absorption in the Schumann-Runge continuum

TABLE VIII-1. WAVELENGTHS FOR SIGNIFICANT PHOTODISSOCIATION AND PHOTOIONIZATION OF SOME IMPORTANT GASES IN THE UPPER ATMOSPHERE

Species	Photodissociation	Photoionization
N_2	None	$\lambda < 796$ Å
O_2	$\lambda < 2424$ Å	$\lambda < 1026$ Å
	$\lambda < 1750$ Å	
O_3	~ 4500 to 7000 Å	None
	~ 2000 to 3400 Å	
H_2O	$\lambda < 2000$ Å	None
CO_2	$\lambda < 1690$ Å	None
O	---	$\lambda < 911$ Å
N	---	$\lambda < 852$ Å

($\lambda < 1750$ Å) takes place above approximately 100 km and results in a predominance of atomic oxygen above some higher level, perhaps 120 km. Very weak absorption in the Herzberg continuum ($\lambda < 2424$ Å) occurs at lower levels and peaks approximately 30 to 40 km.

The latter dissociation results in the formation of very important, although relatively small, amounts of ozone (as discussed below). Ozone in turn is dissociated in the spectral intervals 3000 Å $< \lambda <$ 3400 Å and 4500 Å $< \lambda <$ 7000 Å all the way down to the surface, and in the interval 2000 Å $< \lambda <$ 3000 Å in the stratosphere and above.

Water vapor and carbon dioxide are dissociated in spectral intervals where O_2 absorption is very strong. CO_2 probably remains undissociated to well above 100 km because of this. Water vapor undergoes photodissociation in the mesosphere by quanta in the windows of the Schumann-Runge band system of O_2, 1750 Å $< \lambda <$ 2000 Å.

Photoionization of O_3, H_2O, and CO_2 plays no significant role in the atmosphere. However, N_2, O_2, O, and N are ionized in important quantities by extreme ultraviolet radiation above approximately 100 km.

Apart from direct dissociation, there are certain types of reactions that also cause dissociation. An important one is dissociative recombination, in which an electron and a molecular ion recombine, and the energy released is sufficient to dissociate the molecule; for example,

$$N_2^+ + e \rightarrow N + N \quad . \tag{1}$$

Another is ion-atom interchange; for example,

$$N_2^+ + O \rightarrow NO^+ + N \tag{2}$$

and

$$O^+ + N_2 \rightarrow NO^+ + N \quad . \tag{3}$$

These examples were chosen purposely to demonstrate how atomic nitrogen may be formed, even in the absence of direct photodissociation of N_2.

Although collisions between particles in the ground state do not typically produce dissociation of tightly bound molecules, this may occur if one of the particles is in an excited state so that the reaction is exothermic or nearly so. A possibly important example is

$$O(^1D) + H_2O \rightarrow 2OH \quad . \tag{4}$$

It has been suggested that this reaction may cause significant dissociation of H_2O in the stratosphere (as discussed below).

D. Oxygen Reactions

Reactions involving only oxygen are a basic part of the chemistry of the upper atmosphere. The important ones were first pointed out by Chapman [VIII-1], and their implications have been studied extensively. With M standing for any particle,

$$O_2 + h\nu \rightarrow O + O \quad , \tag{5}$$

$$O + O_2 + M \rightarrow O_3 + M \quad , \tag{6}$$

$$O + O + M \rightarrow O_2 + M \quad , \tag{7}$$

$$O_3 + O \rightarrow 2O_2 \quad , \tag{8}$$

and

$$O_3 + h\nu \rightarrow O_2 + O \quad . \tag{9}$$

In the upper stratosphere, O_2 dissociation is followed by formation of ozone by reaction (6). Ozone in turn is dissociated as discussed above. The terminating reaction in this scheme is (8). Reaction (7) is relatively unimportant at this altitude.

In the lower stratosphere and troposphere, below approximately 25 km, reaction (5) is no longer important because the dissociating quanta do not penetrate so low. Nevertheless, ozone is present in important quantities because of atmospheric transport processes. Once carried to the lower levels, ozone is highly persistent; photodissociation still takes place but is very closely balanced by the reformation of ozone by reaction (6). At these densities, reaction (6) is much more rapid than reaction (8). The ultimate sink for ozone is believed to be near the ground, although other possibilities have been suggested recently.

In the mesophere ozone is still present, but atomic oxygen becomes relatively more abundant since reaction (6) is much slower at the lower density. Reaction (7) becomes an important one. At still higher altitudes, in the thermosphere, the only important reactions in this group are (5) and (7). Because reaction (7) is very slow at low densities (and also because radiative recombination is very slow), oxygen would be almost completely in atomic form above 110 km in a photochemical equilibrium. In fact, there are still O_2 molecules present at higher elevations because of upward transport as a result of mixing and diffusion.

E. Hydrogen-Oxygen Chemistry

Bates and Nicolet [VIII-2] first discussed the very important role of hydrogen reactions in the mesosphere. The principal source of hydrogen is water vapor transported upward from the earth's surface. Once this is dissociated, a large number of reactions involving atomic hydrogen, molecular hydrogen, hydrogen peroxide, and the hydroxyl and perhydroxyl radicals become possible. For more recent discussions, see References VIII-3 through VIII-5. The reaction scheme is very complicated, and many of the rate coefficients are poorly known.

One important result is that the presence of hydrogen significantly reduces the amount of ozone that would be present in an oxygen-only atmosphere. One important reaction is

$$H + O_3 \rightarrow OH + O_2 \quad . \tag{10}$$

At the lower levels odd oxygen may be depleted by the two reactions,

$$OH + O_3 \rightarrow HO_2 + O_2 \tag{11}$$

$$HO_2 + O \rightarrow OH + O_2 \quad . \tag{12}$$

During the last 5 years, there has been much discussion and controversy about the role of hydrogen-oxygen reactions in the stratosphere. A mechanism for the dissociation of H_2O is needed, photodissociation being unimportant below 60 ± 10 km. Following a suggestion of Hampson [VIII-6], Hunt [VIII-3] made elaborate calculations based on the idea that reaction (4) is an important reaction, since $O(^1D)$ results from reaction (9) at short enough wavelengths. The important question is whether reaction (4) is fast enough to compete with the known rapid quenching of $O(^1D)$. There is now some evidence that Hunt used a rate coefficient that was too high for reaction (4) and a coefficient for the quenching process that was too low.

F. Nitrogen-Oxygen Chemistry

Once N_2 is dissociated by processes such as reactions (1), (2), and (3), another large group of reactions involving nitrogen, oxygen, and nitrogen oxides becomes possible. As with the hydrogen-oxygen reactions, rate coefficients are often poorly known.

The importance of nitrogen-oxygen reactions in the mesosphere and lower thermosphere has been known for a long time [VIII-7]. The source of odd nitrogen is principally in the thermosphere, but the atmosphere transports it to the lower levels. One important geophysical consequence is that photoionization of nitric oxide by solar Lyman-α

radiation is a principal source of electrons in the D-region of the ionosphere, under quiet-sun conditions.

More recently it has been suggested [VIII-8] that nitrogen oxides may be important in controlling ozone concentrations in the stratosphere. Reactions such as

$$NO + O_3 \rightarrow NO_2 + O_2 \qquad (13)$$

and

$$NO_2 + O \rightarrow NO + O_2 \qquad (14)$$

could play the same role as reaction (8) does in the oxygen-only scheme.

The origin of odd nitrogen in this part of the atmosphere is not clearly explained. Some may be transported down from the mesosphere. Bates and Hays [VIII-9] have suggested that nitrous oxide produced by biological action near the earth's surface will be transported upward and photodissociated. At wavelengths that are short enough (not available below 25 to 30 km), the products will be N and NO.

G. Transport Processes

There has been occasion in the above discussion to refer to the importance of transport of atmospheric gases by the atmosphere; i.e., ozone in the lower stratosphere and troposphere, nitrogen oxides in the mesosphere and perhaps stratosphere, and molecular oxygen in the thermosphere. It is clearly established now, even to the satisfaction of aeronomers and chemists, that transport processes and chemical processes must be considered together in the understanding of atmospheric composition. How to do this is not established.

In recognition of this problem there has been a recent trend to include vertical transports of trace constituents in the continuity equations of those constituents. With respect to the mesosphere and lower thermosphere, an example of this is the work of Hesstvedt [VIII-4]. This is attempted, typically, by including a vertical transport term of the form

$$F = -K\rho \frac{\partial r}{\partial z} \quad , \qquad (15)$$

where F is the vertical flux, K is the diffusion coefficient, ρ is the total density, and r is the mixing ratio. This formulation has a long record of utility in studies of molecular diffusion; for example, in geophysics its counterpart with respect to molecular conduction of heat is extremely

useful in thermospheric models. However, its use with respect to eddy diffusion or eddy conduction or eddy viscosity, while better than nothing, is dangerous.

From a superficial point of view, without consideration of the physics involved, the difference is that K for molecular transports of various kinds takes on relatively stable and predictable values; one can make a reliable estimate of K and, knowing $\partial r/\partial z$, compute F. For eddy diffusion, K is known to be dependent on properties of the flow, such as wind shear and hydrostatic stability. Its value may be highly variable. This variability can lead to misinterpretations.

Perhaps an example, familiar to meteorologists, will illustrate the point better than general statements. The vertical eddy flux of sensible heat in this formulation is proportional to the negative of the vertical gradient of potential temperature, θ. Near the ground, on the average, $\partial\theta/\partial z$ is positive, so we might conclude that the average flux is downward. In fact, we know from other considerations that the average flux is upward, implying a negative value for K. We can avoid this, however, by realizing that K is very much larger when $\partial\theta/\partial z$ is negative than when it is positive, so that the average value of the product $\partial\theta/\partial z$ is nothing like the product of the average values.

In the case of the downward transport of ozone, from a chemical source near 25 to 30 km to a chemical sink near the ground, we would expect, with constant K and a vertically constant flux, to find more or less constant $\partial r/\partial z$. In fact, we find nothing of the sort. To a first approximation, r has a very small vertical gradient in the troposphere and a much larger one in the lower stratosphere, implying in terms of an eddy diffusivity formulation that K is much larger in the troposphere.

And what of transport through the tropopause? Here the concept of Fickian diffusion comes close to breaking down completely. We know from the work of Reed, Danielsen, and others (for example, Reference VIII-10) that this transport takes place selectively and impulsively, through "breaks" in the tropopause surface, along the so-called stable laminae. For example, radioactive debris deposited in the high latitude stratosphere has a much shorter residence time than that deposited in low latitudes, because midlatitude and high latitude winter and spring are favored places and times for such large scale transfer. It is likely that similar considerations apply to transport in the mesopause region.

In the context of upper atmospheric studies, these examples may not be without their lessons. The argument of Johnson and Wilkins [VIII-11] that K cannot exceed about 10^6 cm^2 sec^{-1} in the mesosphere and lower thermosphere results from dividing the total absorbed heat energy above a level by the average value of $\partial\theta/\partial z$ at the level. This argument does not convince me that K might not be much smaller at

certain times and places, and conversely much larger at certain times and places; for example, in the midlatitude and high latitude winter mesosphere if there are large scale transports similar to those in the tropopause region.

In addition to these problems associated with vertical transport, the problem of horizontal transport poses even more severe problems. For the troposphere and lower stratosphere, north-south transports of heat, momentum, energy, and trace constituents outside low latitudes are accomplished in large part by synoptic-scale eddies; i.e., troughs and ridges of the size that appear on weather maps. These transports may be parameterized in a formal sense by a Fickian-type equation, but the value of K that is appropriate for a particular situation is quite unpredictable and may even be negative. There is not time here to justify or explain these statements in any detail; a useful reference is Reference VIII-12.

Rather than to close on such a negative note, I would like to suggest first that a Fickian transport term is usually better than no transport at all, and secondly, that we are probably going to have to resort eventually to numerical models of the upper atmosphere, models of the type that have been used so successfully in studies of the lower atmosphere. These involve numerical integration of the equations representing conservation of momentum, conservation of mass, and conservation of energy, with boundary conditions and simplifications appropriate to the problem at hand.

For the upper atmosphere, especially above 30 km, the problem will be complicated by the nonlinear interactions between composition and dynamics. The heating term in the energy equation will involve the densities of certain trace constituents that are radiatively active; for example, ozone. To complete the set of equations, continuity equations for these constituents will be necessary, including the nonlinear advection terms from which arise the Reynolds fluxes that accomplish the large scale eddy transports. Futhermore, these equations may involve other trace constituents not directly involved in the energy equation; for example, components that affect the chemistry of ozone. Each of these that enters significantly will require its own continuity equation.

The problem is very complex. A complicated dynamical model, with minimum or no chemistry, or a complicated chemical model, with minimum or no dynamics, may by itself tax the resources of a good computer. It is essential that chemists and dynamicists work together to simplify the problem and retain only the really essential ingredients.

H. References

VIII-1. Chapman, S.: A theory of upper-atmospheric ozone. Mem. Roy. Meteorol. Soc., vol. 3, 1930, pp. 103-125.

VIII-2. Bates, D. R.; and Nicolet, M.: The photochemistry of atmospheric water vapor. J. Geophys. Res., vol. 55, 1950, pp. 301-327.

VIII-3. Hunt, B. G.: Photochemistry of ozone in a moist atmosphere. J. Geophys. Res., vol. 71, 1966, pp. 1385-1398.

VIII-4. Hesstvedt, E.: On the effect of vertical eddy transport on atmospheric composition in the mesosphere and lower thermosphere. Geof. Publ., vol. 27, no. 4, 1968.

VIII-5. Nicolet, M: Ozone and hydrogen reactions. Ann. Geophys., vol. 26, 1970, pp. 531-546.

VIII-6. Hampson, J.: Photochemical behavior of the ozone layer. Canadian Armament Research and Development Establishment, T.N. 1627/64, 1964.

VIII-7. Nicolet, M: Nitrogen oxides in the chemosphere. J. Geophys. Res., vol. 70, 1965, pp. 679-689.

VIII-8. Crutzen, P. J.: The influence of nitrogen oxides on the atmospheric ozone content. Quart. J. Roy. Meteorol. Soc., vol. 96, 1970, pp. 320-325.

VIII-9. Bates, D. R.; and Hays, P. B.: Atmospheric nitrous oxide. Planet. Space Sci., vol. 15, 1967, pp. 189-197.

VIII-10. Danielsen, E. F.: The laminar structure of the atmosphere and its relation to the concept of a tropopause. Archiv. Metor., Geoph., Biokl., vol. A 11, 1959, pp. 293-332.

VIII-11. Johnson, F. S.; and Wilkins, E. M.: Thermal upper limit on eddy diffusion in the mesosphere and lower thermosphere. J. Geophys. Res., vol. 70, 1965, pp. 1281-1284.

VIII-12. Starr, V. P.: Physics of Negative Viscosity Phenomena. New York, McGraw-Hill, 1968.

CHAPTER IX. THE FLARE FORECASTING FIASCO – IS EVERYBODY LOOKING AT THE SAME SUN?

By

Frederick W. Ward, Jr.
Space Physics Laboratory
Air Force Cambridge Research Laboratories,
Bedford, Massachusetts

A. Abstract

The passionate dream of the oppressed solar flare forecaster is the scientific discovery of the cause, the energy source, and the observable antecedent conditions for the solar flare. This "tunnel vision" has had a profound and deleterious effect on the state-of-the-art. It has diverted attention from many promising approaches and relegated them to limbo. More importantly, it has actually inhibited the scientific search for the underlying physical mechanisms. Examples of neglected aspects of the problem cover the entire range from observing techniques and equipment to the availability and accuracy of the archived scientific data. A cursory inspection of this data leads to an obvious conclusion: there are at least two suns up there.

B. Introduction

I must confess to a proclivity for humor in many of the serious aspects of human organization. The title of this paper could be presumed to be just one more example. However, in the present instance, what might appear to be poking humor is in reality a measure of deep frustration. The events that led to this frustration will, I hope, not only underscore the serious nature of our present shortcomings but also inspire a determination to work toward rectification of the present difficulties.

You might also assume from the title that we are really dealing with only a single problem, when, in fact, the nature of solar activity is such that the problem can be easily and reasonably subdivided into at least five distinct parts, each somewhat independent of the others. These five areas are as follows:

1. Forecasting a general level of solar activity.

2. Forecasting the development and decay of particular active regions.

3. Forecasting the beginning of a particular flare.

4. Forecasting the course of development of a flare that has just started.

5. Forecasting the radiation spectrum of the flare, otherwise known as forecasting its geophysical effects.

Examples of the types of forecasts falling in these five categories show the extent of our knowledge rather directly. Forecasts of the general level of activity are couched in the rather loose form of "increasing," "decreasing," or "little change." Forecasts of the development and decay of a particular active region are phrased in terms of the probability of a flare within a specified period of time in the future. Forecasts of the beginning time of a flare are stated in precise terms, usually minutes, but are the subject of considerable debate among the forecasters themselves. Forecasts of the course of development of a flare, once it has started, are in crude class categories and are subject to verification by the forecaster himself. Finally, forecasts of its radiation spectrum are noted for extreme overkill, and are generally of a yes/no type.

Type of Forecast	Specificity	State-of-the-Art
1. General Activity Level	Increase/Decrease	Almost No Skill
2. Development of Region	Probability (%)	Reasonably Good Skill
3. Beginning of Flare	Minutes	Highly Debatable
4. Development of Flare	Classes	Highly Debatable
5. Geophysical Effects	Yes/No	Some Skill

For any readers who might be strangers to this subject, I think it would be helpful if I were to give you a general feeling of what the present state-of-the-art is in a reference frame that should be quite familiar. Imagine, if you can, an office in a large company populated predominantly by males. Now, introduce a shy secretary, married approximately five months. The water-cooler cowboys have undoubtedly wondered from time to time how much longer our newly married secretary will continue to work. This can be phrased more succinctly by asking the question, "Is she pregnant yet?" The quasi-scientific approach would be to look for one or more of the following symptoms: late arrival, an early morning "hang dog" look, otherwise known as morning sickness, or a protruding tummy. The more mature, experienced males will also try to gauge the extent of enhanced voluptuousness. This rather amateur approach to scientific inquiry contrasts markedly with the approach of the objective, trained obstetrician. These wise gentlemen are more prone to require tests that have proven to yield rather accurate results.

The forecasting of solar flares bears a striking resemblance to water-cooler science. There are serious claims by well qualified individuals that a trained observer looking at a particular active region on the sun can tell whether that region is likely to produce a flare in the very near future. In the same way that the water-cooler observations yield occasional correct forecasts, the subjective solar forecaster occasionally correctly forecasts the onset of a solar flare. The types of phenomena that this trained observer watches for are configurations of the spots and the magnetic fields, dark filaments that wink, and a very steep gradient in the radio brightness temperature.

Another technique that has many adherents avoids all of these difficulties. I will mention it only this once; even that is more attention than it warrants. This technique depends on the relatively straightforward forecasting of the positions of the planets around the sun. Depending on the particular investigator, the development of active regions and the occurrence of flares depend on some combination of gravitational forces, the acceleration of the sun around the center of the solar system, a change of the acceleration of the sun around the center of the solar system, or the change of the change of acceleration of the sun around the center of the solar system. All these methods have one feature in common, namely that each yields an apparent relationship between the development of active regions and the predictor variables. Such an apparent relationship is absolutely guaranteed because of the relatively limited degrees of freedom in the predictor data. Careful verification of the forecasts made by these methods yields the expected results; that is, there is serious doubt that these forecasts show any skill.

C. The Development of the Active Region, the Flare, and Its Effects

For those readers who may have very little knowledge of what is considered the "typical" active region and its morphology of development and decay, this review may be somewhat too brief, for which the later discussion should be of some assistance. I refer to this morphology as a "typical" morphology of active region development because that is a common belief. Nothing, however, could be further from the truth. The word "typical" has been used because such development leads to exciting and interesting happenings for the solar astronomer. It is very important to keep in mind that the typical (without quotes) active region never becomes very active, grows only to a small size, and fades away with hardly a memory.

The untypical "typical" active region starts as a small brightening observable in the light of the emission lines of certain elements, most commonly calcium or hydrogen. It can also be observed in white light, but only when near the edge of the disk. In white light this brightening is

called a facula; in the emission lines it is called a plage. For our "typical" active region, the plage (facula) grows over a period of weeks to cover an area of 1 percent or more of the solar disk. A few days after the first brightening is noticed, the first spots are seen. These are usually very small spots and these (and other) spots grow over the course of 1 week into a large sunspot group. More often than not, this group is divided into two groups of spots; a large spot with or without companions near the leading (western) edge of the group, and another large spot with or without others toward the following (eastern) edge of the group, with the group aligned in a more or less east-to-west direction. In the course of 1 or 2 weeks the group grows in size to cover an area only a small fraction of 1 percent of the solar disk, surrounded by the much larger plage. During the development stage of a group, as the group is increasing in size, there are short-lived enhancements of radiation, the largest of which may cover an area only a fraction of 1 percent of the solar disk and most are less than one-tenth of 1 percent of the disk area. These enhancements are short-lived, ranging from some minutes up to a few hours, the larger enhancements usually lasting for longer periods of time. Measurements in the radio-frequency spectrum of the temperature of the solar atmosphere in the chromosphere and corona above the active regions show very hot spots above the developing active region. The temperature in these hot spots tends to increase during the very active stages of the development of the region. After an activity period of some few weeks, the region tends to decay in a roughly reverse manner with the radio brightness temperature decreasing, the sunspot size decreasing, and the structure becoming much less complex. The frequency of flares decreases, and the plage size also decreases, but more slowly. After a period of 1 month or more the region becomes quiescent, the spots gradually die out, and after another month or two the plage itself also becomes indistinguishable. The frequency of development of active regions follows the sunspot cycle, increasing and decreasing over the course of 10 to 11 years.

All of these things have been known for many years and have provided the method by which forecasts of flares can be and have been made. The general level of solar activity is forecast from projections of the sunspot cycle and an introspective analysis of the state of one's digestion. This type of forecasting is closely akin to long range weather forecasting and has attracted a similar type of person. Little of this paper will be devoted to it. Forecasting the development and decay of particular active regions has proven a fruitful area of study because of many observational keys available, and this will occupy the bulk of this discussion. Forecasting the beginning of a particular flare has been left primarily to the observer at the telescope monitoring the small scale, short-lived phenomena within the active region. After the onset of the flare, the forecasting of its development and its radiation spectrum become extremely important and are likely to be the two areas of most concentrated effort in the coming years. To date, both areas have received much less attention than they warrant.

Before proceeding further, there are two complicating facts of life that require emphasis. The first fact is that a particular sunspot group is constantly changing. There are many spots breaking out and decaying at the same time; the radio brightness temperatures in the chromosphere and corona are constantly changing, the plage area surrounding the spots is in constant change, and the rate at which flares occur varies from one day to the next. Everything about the region is in constant change, and the changes are not completely regular. However, all of the features that we use to describe the change are partially interrelated. When we speak of a region becoming more active, we mean in general that the spot group, the plage, the amount of flariness, and the radio brightness temperature are all increasing in size and/or complexity. For a particular active region, however, it may be that only the radio brightness temperature is increasing at any particular time or that the amount of flariness has suddenly increased, or that the spot group is rapidly enlarging and becoming more complex. All active regions do not follow the same course.

The second complicating fact is that the development of a particular active region is completely independent of the development or decay of all other active regions on the surface of the sun at the same time.

D. Forecasting Problems

In a sense the title of this section is somewhat presumptuous. I suspect that if the total range of problems in this area could be identified, some better solutions almost surely would result. However, there are some obvious categories of severe problems, and these run the gamut from the original observations through the handling of the data, the forecasting methods, and the verification or lack of verification of these forecasts.

I would like to begin with the forecast verification, because I think this best illustrates the depths of our ignorance. As an aside, consider the following hypothetical situation. Assume that we know the physics of flare development completely! Further, assume that there are useful precursors of the flare and that they are observed accurately. The forecast of the flare is therefore made — correctly, 100 percent correct; but, the verifying observation of the flare was either missing or in error! In such a situation we would never know that a perfect forecast was indeed made. Our records would show either a wrong forecast or an unverified one. This situation, it should be noted, is substantially superior to the present one and points up the absolute necessity of a total scheme starting with good observations and extending to and through a comprehensive forecast verification. In our present imperfect world, the feedback of improvements from one area to another is even more noticeable.

In discussing flare forecast verification, certain points are well accepted. For example, most everyone agrees that the forecasts are not perfect and with the exception of extreme cynics, everyone agrees that the forecasts are better than pure chance. Unfortunately, the range in between these limits is rather extreme, and the accuracy achievable within the present state-of-the art is a subject of considerable debate. There has been little in the way of verification statistics, and this vacuum has proven a very fertile ground for the growth of unsubstantiated claims. It should be obvious that you cannot know whether you are doing better than someone else, or better than what you did before by some new or improved or modified technique, if you did not know how well you were doing before. This might seem an elementary principle; that it has been ignored is astounding!

This avoidance of the hard realities of verification has its price directly in forecast accuracy, as the following reasoning may show. Given a scheme A that yields a forecast of parameter X with a reliability R; and given another scheme B, which also yields a forecast of X, with a reliability of S; and further, given that scheme A and scheme B are somewhat independent of each other; then there exists a scheme C, which is a combination of scheme A and scheme B, that will yield a forecast of parameter X, with a reliability T, where T is larger than either R or S. It would seem reasonable under the present circumstances, with many hundreds of schemes, at least some of which must be partially independent of one another, that there must exist a consensus scheme that is better than any one of the schemes presently in existence. A verification procedure, a well designed verification procedure, would surely yield some rather interesting results and better forecasts. The results would be useful for more than just a comparison of different techniques or different forecasters. They would yield some rather interesting insight into the physical processes of the flare mechanism.

This brings us to the problems of forecasting. Many of you have been involved in meteorology and have either participated in, or been observers to, what are euphemistically called "map discussions." I can recall a map discussion many years back in which an extremely competent meteorologist spent about 10 minutes in elaborate self-praise of his forecast from yesterday. As I recall his comments spanned the range from "notice the position of the front moving down through Florida" and "the development of the system in the Gulf was in the right direction and the timing was rather good," and "I even picked up the ridging along the west coast." It was during the general comments by the audience that someone noticed that the forecast map was, in fact, from last week, not from yesterday. Despite a few snickers, and without missing a stroke, this gentleman proceeded to verify the right forecast map in almost the same terms, and with similar claims of accuracy. The fact that the two forecasts were quite different mattered hardly a whit.

CHAPTER IX
FREDERICK W. WARD, JR.

Meteorology in the last 20 years has at least come to the point where a large fraction of the forecasting information is generated by computers. I certainly make no claim that computers have any more brains than human forecasters, but the use of computers has had one very beneficial effect. It forecasts by a known technique and uses that technique for an extended period of time. When a change of the forecasting procedure is made, it necessarily involves a rather large amount of reprogramming effort. The magnitude of this effort is sufficient so that considerable comparative testing is done before any new computer program is put into operational use. This yields not only the comparison between old and new, but can, if verified by a technique that is consistent over the years, yield a rather interesting overall comparison with all techniques used in the past and give a measure of long term forecasting improvement.

We have not as yet achieved a similar situation in the field of solar flare forecasting. Flare forecasting is now done by a number of techniques. There are, for example, many experts in the field who have a "sixth sense" about which regions on the sun are likely to produce flares and which are not. They have a "seventh sense" which tells them when these particular regions are going to flare. By contrast, there are a number of objective forecast aids available. These are generally statistical tables or equations that give a probability of a particular region yielding a flare within a certain period of time in the future. In addition to these purely subjective and objective techniques, there are people who start with the tables and the equations and then incorporate their own subjective judgment. Finally, there are forecasters who start with someone else's forecast, add the information available in the tables and equations, and leaven all this with their own subjective feeling of the situation to arrive at their forecast. If more than one forecaster is on duty at some point in time, there will likely be more than one forecast.

Part of the reason for the different forecasts can be traced to the difficulties in determining which pieces of data apply to which active regions. As noted earlier, each active region pursues a development history that is independent of all other active regions on the sun at the same time, and apparently at all other times. Therefore, if a flare occurs, for example, at 11:05 on any particular day, and there are three or four active regions on the sun on that day, it is absolutely necessary that the forecaster know in which active region the flare has been seen. It would seem a simple problem to identify the correct active region, since the observer recording the data can see all the active regions at the same time. However, the ways of solar observers are strange indeed, and many gremlins afflict the recording and communication of the observations. Back at the forecast center, the duty forecaster occasionally finds himself with a flare that apparently occurred outside of any region currently active. He then must make a determination as to whether or not the flare would fall in a "juicy" active region by changing the reported sign of the flare latitude. If not, he looks to see if there are other reports of the same flare from

another observatory. These additional reports, when available, sometimes help; at other times, they only further confuse the situation. The accuracy of any forecasts that depend on these data is obviously a function of the accuracy of the forecaster's matching of the flare reports with other data concerning the active region.

We tried some experiments many years back to try to determine what rules might be used by an experienced solar forecaster to match or collate all of the available solar data. The idea was to try to build this experience into a computing machine program that would perform this rather laborious task in real time. To this day we have not found a simple set of rules to which all trained solar observers subscribe. It is simply not possible at present to unequivocally associate all the available solar observations with the proper active region. This is true even when the person doing the collating is allowed unlimited time and access to all of the original observatory records and films. From this it follows that the history of active regions must be imperfect, and such imperfections will degrade the accuracy of forecasts.

This brings us back to our basic source of difficulty, the observer himself. The observer's life is not exactly pleasant. Under the guise of training, most of the observers have been thoroughly brainwashed by other experienced observers into believing in certain articles of faith. In addition, there have been many well-intentioned people over the years who have tried to make the observer's life somewhat simpler and to ensure that the observations were both accurate and consistent. Toward this end, many years of effort have been directed into making international observations directly intercomparable. Much of this effort has gone into designing the procedures or the rules by which observations are to be made, the formats by which they are to be reported, and the procedures to be followed in reporting them. It was natural to assume that not too much effort would be necessary in accumulating and utilizing these observations to which much apparent care had been devoted. At the Air Force Cambridge Research Laboratories we decided to accumulate the data for one full sunspot extending from the sunspot minimum in 1955 to the following sunspot minimum in 1964. This period included the international geophysical years of 1957 and 1958, around sunspot maximum. Lots of effort was expended in the accumulation and checking of these data. One of the important conclusions we drew from all this effort was that some of our most serious problems were a direct result of the care devoted to the international organization of the observations. If this seems surprising, one example might serve as illustration.

The International Astronomical Union (IAU) had set up rules for making observations of solar flares. However, in the name of science these rules were constantly being modified. Observations made in 1955 did not conform to those made in 1958, and neither conformed to those of 1964. This would have been sufficiently troublesome except for the fact that when the rules changed not all observatories changed at the

same time. So that at any one point in time two or three different set of rules were being followed simultaneously. A few observatories knew "better" and followed their own private rules, which, of course, changed as observers changed. The result was chaos. The sincere attempts to produce useful and consistent data were in fact working in just the opposite direction.

There were more than 50 observatories around the world making observations of solar flares during that decade, and there were lots of reports. We had close to 100 000 flare reports in the 10 years. If reporting had been up to current standards, there probably would have been twice that number. But the data from different observatories did not match. The mismatch was so bad in many cases that one wondered which aspects of the observations were wrong. Was the time of day in error? Or was it the day of the month, or the month, or even the year? Or were the measurements themselves grossly in error? Yet with few exceptions every observatory sincerely claimed that its data were right. We spent a couple of years in meetings and discussions with experienced solar observers without any real agreement on what was producing the differences. We must have had at least 100 possibilities for sources of difficulties. We even tried an experiment in which we sent out a picture and a drawing of a sunspot and asked a number of observatories around the world to make measurements of the size of the sunspot. The results were unbelievable!

To summarize the problem, it is reasonable to say that if there was a way in which a faulty technique could be incorporated, it was! There are also basic difficulties in the observational equipment. There are fundamental difficulties in the collation of the data to form the history of active regions. The present forecasting techniques suffer from the severe limitations of fundamental knowledge. Lastly, the lack of a comprehensive verification program is nothing short of criminal.

E. Solutions

The ultimate answer to the problem of forecasting solar flares lies, of course, in the complete understanding of the physical processes that produce the flare and an ability to observe the requisite antecedent traditions. That this is not a likely possibility in the near future can be seen from many of my preceding comments. First, the observations are not sufficiently accurate and complete so that we always know when a flare has occurred. Even in those cases where there is agreement that a flare has occurred, there is a wide disparity in the measurement of its size, its brightness, the time when the flare began, and other parameters that describe its general development and decay. Additionally, if one of the necessary precursors turns out to involve changes in a parameter such as the magnetic field structure of the sunspot region, then the forecasting accuracy will be limited because the measurements of this

phenomenon are incomplete and imprecise. In the ultimate, of course, it may be that the really important part of the process will be described by a phenomenon that is not even observed or measured at the present time.

Since a total understanding of the physics and dynamics of the flare is an unlikely possibility in the near future, we must necessarily turn to statistical studies of the flare in relation to other useful precursors. We can take advantage of a well-known and fundamental property of a statistical ensemble of data; namely, that the error in the estimation of a mean value decreases as the number of observations going into the estimate increases (the error is inversely proportional to the square root of the number of observations). However, there is one basic assumption that must be fulfilled for the error in the estimation of the mean value to decrease as more observations are accumulated. This assumption is that the observational estimates themselves are unbiased estimates of the true value. What this means in practice is that observations of a particular flare from a number of different observatories will yield a more accurate estimate of the flare's properties than could be obtained from a single observation, provided that all observatories contributing to the estimate made measurements by similar instruments and techniques. If, on the other hand, each observatory made its measurement by a basically different instrument and/or technique, then the estimate of the flare properties made by combining the estimates from each of the observatories might have characteristically larger errors than those from some of the individual sites. To say that the solution to the problem lies in more and "better" observations is really much too simple.

There is as yet, no unanimity on what would compose "better" observations. Many qualified observers feel that there should be emphasis on new and different measurements, while others want measurements in more detail. There is strong support at present for data on the sunspot structure and its magnetic field configurations in considerable detail. Others quite reasonably feel that there should be less emphasis on the ground-based observations and more emphasis on observations from satellites, which can provide data on the X-ray spectrum of the radiation. Still others feel that more detail in time would considerably enhance the present observations. The point is that "better observations" mean different things to different people.

I would like to suggest an approach, which has the advantage that if it does not lead to an acceptable improvement, it at least establishes our present capabilities precisely and eliminates the major fraction of our present differences of opinion. This approach emphasizes the areas where properly phrased questions and properly designed experiments are guaranteed to yield useful (albeit possibly negative) answers. The approach scorns the "cocktail science" approach, in which the broadcasting of untested ideas is rampant. This field does not lack for ideas; the dearth is in tested and testable hypotheses. There are dozens of

ardent adherents, but few objective judges. One could say, with some considerable justification, that the field has a peculiar appeal for undisciplined scientists.

With that for an introduction, I will describe what is basically the Air Force approach to the problem of improving our capability to forecast solar activity. I take some pride of authorship in this program, having seen it through its birth pangs and fought hard for some years to nurse it and wean it. I would not want to give the impression that there is not some disagreement from other Air Force people involved in the program. These disagreements are minor, however, compared with the differences of opinion from outsiders. I believe, however, that the Air Force has a rather distinct advantage in this field: It has few past actions to justify, either legal or moral.

In contrast to the past efforts of other organizations, the Air Force program puts great emphasis on radically improved observations and analysis. A corollary decision was also made to press for the use of objective analysis techniques and computers to handle the ever-increasing volume of data being collected and processed. The decision that I feel will yield the most important long term benefits was to use these objective techniques and computers to collect, correct, collate, analyze, and prepare all these data for easy and efficient study and analysis by the scientific community. The data will be available to the scientific community before it is 2 to 3 months old. The quality and quantity of the data and the care and speed of the analysis will surely yield some substantial scientific dividends in the coming years.

First, let us look at a couple of the improvements in observing equipment and techniques. We considered conducting an extended experiment to try to determine why two solar observers, presumably looking at the same sun, reported measurements that were most easily interpretable as coming from two different suns. After much discussion, we accumulated dozens of possible reasons for the large discrepancies. These possibilities ranged from "seeing" difficulties to human contrariness. Their sheer number and diversity forced us to a totally different approach.

This approach consisted basically of holding as many variables constant as was reasonably possible and hoping that the results would eliminate our double vision. Figure IX-1 is a picture of the result, the first pair of matched solar telescopes. These are slightly modified Spectrolab telescopes. The modifications consist of a very careful matching of optical characteristics. The filters were specially selected for matching characteristics, the intermatch being more critical than the actual bandwidth or position around H-alpha. The precise details of the observations were of secondary interest to the question of intertelescopic consistency. As it turned out, the filter width and the position of

the filters with respect to the line center came quite close to those we desired anyway.

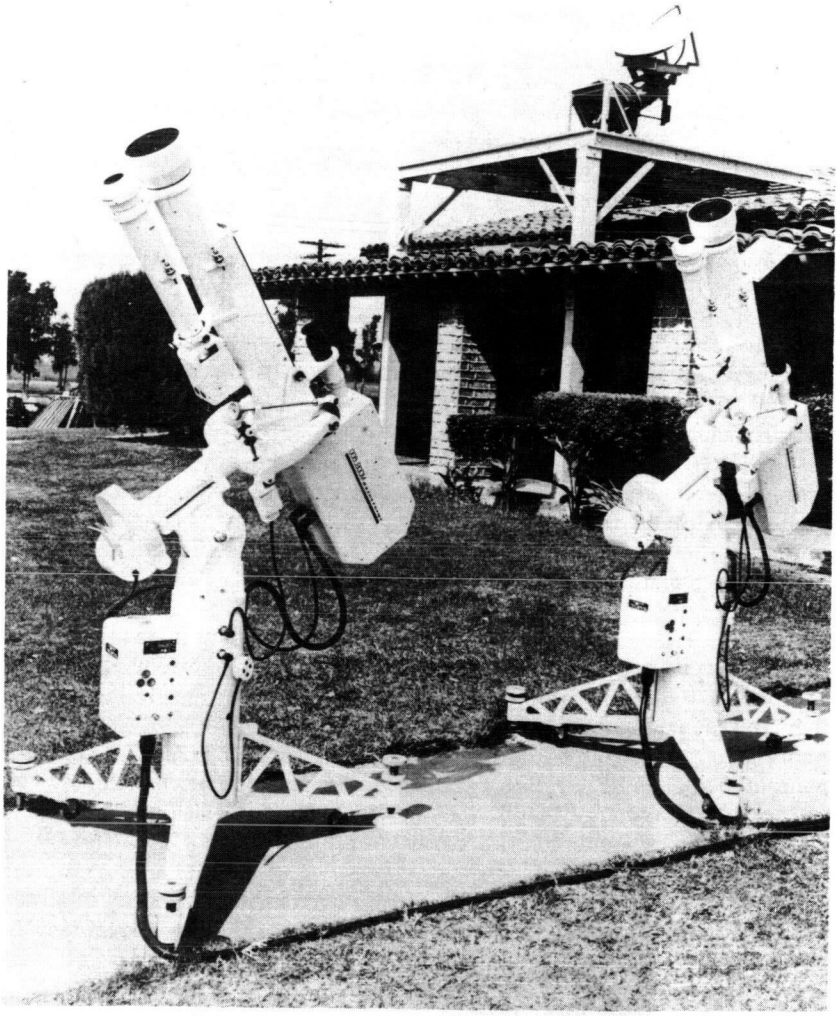


Figure IX-1. Matched pair of solar telescopes.

To reduce the influence of unpredictable observer behavior, this pair of telescopes is equipped with videometers, also with matched characteristics. The videometer is basically a television camera/computer combination. The camera scans either the entire solar disk or preselected areas. The computer selects and analyzes the appropriate data

and delivers the results in either an analog or digital form. We are currently programming our mini-computers to deliver the flare area, peak brightness, and integrated light surplus once every 5 to 10 sec during the life of each flare. The camera is planned at the moment to scan up to four (preselected) active regions simultaneously. The ultimate arrangement for active region selection, scan cycle, and data to be analyzed and printed out is yet to be finalized, and will depend on the results of a forthcoming test program and some operating experience.

This test program will be extensive. We will emphasize primarily the consistency of observations between the two telescopes. We plan to demonstrate this consistency first in a side-by-side arrangement at our test site in Corona, California. The side-by-side configuration should obviate any inconsistencies resulting from seeing differences. If these results are encouraging, one scope will be moved to the Air Weather Service (AWS) operating location at Upper Van Norman reservoir, about 50 or more miles to the north and west. The testing will be continued under this minimum separation until all the obvious bugs are out of both the equipment and the procedures. Modifications can easily be made to the equipment at this point, if necessary. Final testing will commence with the installation of one telescope at an AWS operating location well removed from Corona (approximately 1000 miles) into an area where the meteorological effects are totally different.

If the telescopes still yield consistent measurements at this point, the main purpose of the test program will have been successful; we will have demonstrated that there is really only one sun!

Another important part of this Air Force program is almost completed. This pertains to the development and implementation of new reporting codes. New coding rules and formats have been worked out that are designed not only to minimize the number of errors resulting from observer malfunction, but also to allow easy and efficient decoding, correction, and use of the observational data by digital computers. The older coding rules were designed to make extremely efficient use of teletype time, at the sacrifice of both observer time and accuracy. Our tests showed conclusively that this was a very bad compromise. That situation has now been essentially corrected. The new coding rules and reporting formats make somewhat less efficient use of the teletype time, but they drastically reduce the errors.

By far, the largest fraction of our energy has been devoted to reducing and analyzing the data to be received by the central forecasting facility. The major tasks were editing, collating, and analyzing, and the largest and most complex single computing program was the one that constructs the active region history. This program searched through all the plage, spot, flare, radio, and magnetic data and assigned them to the appropriate active region. The problems that were solved in the process of working out the rules for this program have haunted investigators in

this field for two decades. One problem involved the assignment of each flare to its appropriate spot, a task of very considerable magnitude, especially when one considers that the ultimate judge of all disputes is not easily contacted. Other problems arose because there are regular differences of opinion over the limits of the spot group itself and which spots are part of which group. A measure of this disagreement can be gleaned from the disparity of spot group numbering noted in the Greenwich Photoheliographic results between Greenwich and Mt. Wilson, two of the most careful measures available.

The consequences of a basic difference between the Mt. Wilson magnetic class data for a single group that Greenwich has classified into two groups, or vice versa, are easy to imagine. For the analysis of past data, one can simply ignore the cases in which discrepancies occur; the forecaster in real time has no such luxury. I personally feel the answer to this type of problem lies straightforwardly in arbitrary and binding group assignments by the forecast center and in a requirement that all observing sites comply. Such a radical realtering of assignments will, however, not be easily accomplished.

In the business of forecasting itself, our approach has been straightforward, but limited. We are trying to filter the maximum amount of information out of the available data by judicious application of the best statistical techniques. We have limited our efforts to generating a number of aids for the forecaster. The reason for this limitation lies in the nature of the problem. No physical theory is available, but in the vacuum, the quantity and diversity of data on parameters that are possible predictors have been rapidly increasing. Statistical techniques require a historical sample of data of reasonable size. The forecaster should not necessarily wait for the acquisition of a large sample of data. His job requires some on-the-spot experimentation with quite limited samples. Therefore, we have devoted our effort to filtering information out of only those data for which a consistent record has existed for at least a couple of years. We provide the forecaster with a preliminary (or base) forecast which has within it all the information obtainable from specified parameters only. He can then modify the basic forecast using only that data not already incorporated. This procedure has not only the obvious advantage of more effective and efficient use of the forecaster' s time, but is is bound to yield a forecast of an accuracy as good as or better than the forecaster could achieve on his own. It has the unique advantage, in addition, of allowing a rapid determination of the value of new data sources merely by a comparison of the accuracy of the basic forecast with the forecaster' s ultimate skill using all available data.

Less emphasis to date has been applied to the development of a comprehensive verification program. This neglect can be excused because of the precedence of other problems and the lack of easy access to a large computer at the present time. We fully expect, however, that increasing emphasis will be devoted to this subject over the next few

years as the Air Weather Service automated data-handling system becomes operational.

A very important aspect of the Air Force Program is quite simply to encourage the scientific community at large to devote increased effort to this problem. We hope to achieve this goal by making the scientist' s life easier. We plan to make the analyzed data from the central processing facility available in convenient form and reasonably error-free. Our experience of the last 6 years in working with this data has shown that it is both difficult and frustrating to work with the data in its present form. The errors, inhomogeneities, and biases result in much tedious and time-consuming effort. It is almost as easy to go out and re-do the observations themselves. The effort that the Air Weather Service is now expending on observational consistency and reporting, and will spend in the future on error checking and processing, will drastically alter this situation. We plan to make this data and its unique characteristics well known among members of the scientific community interested in problems of solar flares, and their causes and effects.

F. Summary

In summary, I would like to restate my basic point — the best way to improve the accuracy of solar forecasts is not necessarily by trying to improve the forecasting techniques directly. The overall accuracy and usefulness of the forecasts are determined by many factors, only some of which can be manipulated at any particular time. In the domain of forecasting solar events and their effects on the earth, we found that the subjective forecasting techniques already in use were not so bad and that some simple modifications would raise them to temporarily acceptable levels. By contrast, many other factors directly influencing the forecast accuracy in a major and deleterious way needed drastic revision. Changes were necessary in the basic observational equipment, in the procedures for recording and transmitting the data, in the processing and reduction of the data, in analysis, forecasting and verification, and in the treatment and preparation of the data for archiving. Every one of these factors directly contributes either to the present forecast accuracy or to the groundwork for future improvements.

It has been said many times that a good observation is the best short range forecast, and in too many instances, this is true. The reverse is even more true; that is, a bad observation is not only a bad short range forecast, but it generates bad long range forecasts and absolutely guarantees that the forecasts will not improve.